

# Open Research Online

---

The Open University's repository of research publications and other research outputs

## Information enforcement in learning with graphics : improving syllogistic reasoning skills

### Thesis

#### How to cite:

Dobson, Michael W. (2000). Information enforcement in learning with graphics : improving syllogistic reasoning skills. PhD thesis The Open University.

For guidance on citations see [FAQs](#).

© 2000 The Author



<https://creativecommons.org/licenses/by-nc-nd/4.0/>

Version: Version of Record

Link(s) to article on publisher's website:

<http://dx.doi.org/doi:10.21954/ou.ro.0000dc92>

---

Copyright and Moral Rights for the articles on this site are retained by the individual authors and/or other copyright owners. For more information on Open Research Online's data [policy](#) on reuse of materials please consult the policies page.

---

[oro.open.ac.uk](http://oro.open.ac.uk)

UNRESTRICTED

# INFORMATION ENFORCEMENT IN LEARNING WITH GRAPHICS: IMPROVING SYLLOGISTIC REASONING SKILLS

BY

MICHAEL WILLIAM DOBSON  
BA (HONS) MANCHESTER  
MSc SUSSEX

A THESIS OFFERED FOR THE DEGREE OF DOCTOR OF PHILOSOPHY IN  
EDUCATIONAL TECHNOLOGY

INSTITUTE OF EDUCATIONAL TECHNOLOGY  
THE OPEN UNIVERSITY  
MILTON KEYNES

AUTHOR'S NO: M70883SX  
DATE OF SUBMISSION: 26-JANUARY-1999  
DATE OF AWARD: 30-MAY-2000

JANUARY 1999

<b>ACKNOWLEDGMENTS</b>	<b>VII</b>
<b>ABSTRACT</b>	<b>IX</b>
<b>1 INTRODUCTION &amp; OVERVIEW</b>	<b>1</b>
1.1 MODALITY ASSIGNMENT IN TEACHING WITH GRAPHICS.	1
<b>2. LEARNING WITH DIAGRAMS</b>	<b>11</b>
2.1 DEFINING GRAPHICAL REPRESENTATIONS.	12
2.2 DIAGRAM PROCESSING.	19
2.3 TRANSLATING BETWEEN MULTIPLE REPRESENTATIONS.	22
2.4 DISCUSSION & IMPLICATIONS FOR FURTHER STUDY.	23
<b>3 REPRESENTATIONS &amp; BIASES IN HUMAN REASONING</b>	<b>27</b>
3.1 SYLLOGISTIC REASONING.	27
3.2 HUMAN PERFORMANCE.	29
3.3 REPRESENTATIONS OF SYLLOGISTIC REASONING.	36
3.4 IMPLICATIONS FOR STUDY.	56
3.5 METHODOLOGY.	58
<b>4 SPECIFICITY APPLIED</b>	<b>61</b>
4.1 QUANTIFYING SPECIFICITY.	61
4.2 THE OFFICE INDICATOR SYSTEM.	62
4.3 ANALYSIS OF VENN CIRCLE SYSTEM.	65
4.4 THE EDINBURGH-EULER SYSTEM.	77
4.5 CARROLL'S 'GAME OF LOGIC.'	86
4.6 SUMMARY.	88
<b>5 SOFTWARE &amp; PILOT STUDY</b>	<b>93</b>
5.1 TRIALING SUBJECTS AND MATERIALS.	93
5.2 DIFFERENCES AMONG SOFTWARE.	100
5.3 THE STUDY.	102
5.4 IMPLICATIONS FOR FURTHER STUDY.	108
5.5 SUMMARY.	109
<b>6 EVALUATION OF THE SPECIFICITY PRINCIPLE</b>	<b>111</b>
6.1 MOTIVATION.	111
6.2 EXPERIMENTAL DESIGN.	111
6.3 RESULTS.	116
6.4 DISCUSSION.	138
6.5 DESIGN IMPLICATIONS	141
<b>7 PRIOR EXPECTATIONS &amp; LEARNER STYLES</b>	<b>143</b>
7.1 INTRODUCTION.	143
7.2 COGNITIVE STYLES.	145
7.3 THE CARD SORTING TASK.	151
7.4 TRAINING CASE STUDIES & TEMPORAL ANALYSIS.	168
7.5 DISCUSSION.	180
<b>8 DISCUSSION &amp; CONCLUSIONS</b>	<b>183</b>
8.1 CONTRIBUTIONS.	184
8.2 LIMITATIONS.	194
8.3 FURTHER WORK.	196

8.4 IMPLICATIONS FOR EDUCATIONAL TECHNOLOGY.	210
REFERENCES	213
APPENDICES	227
DATA SET FOR SPY RING TEST.	228
INTRODUCTION TO ARISTOTELIAN LOGIC: THE SYLLOGISM.	229
USING Tarski's World ™ TO LEARN ARISTOTELIAN LOGIC	231
THE EULER CIRCLE SYSTEM AND THE SYLLOGISM.	239
THE Venn SYSTEM AND ARISTOTELIAN LOGIC.	244
PRE-TEST	248
CODE FOR THE Venn COUNTING ARGUMENT	252



TABLE OF FIGURES

FIGURE 1: QUALITATIVE STAGES OF THOUGHT..... 14

FIGURE 2: A MODEL OF VISUAL REASONING ..... 21

FIGURE 3: WASON'S DIAGRAMS..... 33

FIGURE 4: CONDITIONS FOR SOME X ARE Y. .... 34

FIGURE 5: VENN & EULER ILLUSTRATIONS OF THE FOUR CONTINGENT STATES WHICH MAY BE TRUE OF  
SOME X ARE Y..... 35

FIGURE 6A AND 6B: VENN SHOWING ALL B ARE A, SOME B ARE C. (6B SHOWS A PART SPECIFIED  
INDIVIDUAL). .... 38

FIGURE 7: THE OLDER EULER METHOD..... 41

FIGURE 8: THE EDINBURGH EULER METHOD..... 43

FIGURE 9: PREDICATE CALCULUS METHOD..... 49

FIGURE 10: THE CARROLL METHOD..... 54

FIGURE 11: OFFICE INDICATOR SYSTEM (ASSUMING ROTATIONAL EQUIVALENCE)..... 63

FIGURE 12: PREMISES IN EDINBURGH EULER. .... 77

FIGURE 13: COUNTING EULER BY 'REACHABILITY'..... 83

FIGURE 14: THE CARROLL SYSTEM ..... 86

FIGURE 15: COUNTING CARROLL'S SYSTEM..... 87

FIGURE 16: EDINBURGH EULER SOFTWARE..... 95

FIGURE 17: VENN SOFTWARE..... 97

FIGURE 18: TARSKI'S WORLD SOFTWARE. .... 98

FIGURE 19: MESSAGES BETWEEN AGENTS AS GRAPHS OVER 3 YEARS. .... 148

FIGURE 20: A GEOGRAPHICAL SOLUTION FOR THE SPY RING..... 149

FIGURE 21: TIMELINES DATA FROM EULER CONDITIONS, (S1 & S2 ARE SUPPORTED WITH THE SOFTWARE,  
S3 & S8, USED PAPER AND PENCIL)..... 174

FIGURE 22: TIMELINES DATA FROM VENN CONDITIONS, (S5 & S6 ARE SUPPORTED WITH THE SOFTWARE,  
S7 & S8 USED PAPER AND PENCIL)..... 175

FIGURE 23: MAP OF STUDENT ERRORS IN SYLLOGISTIC REASONING..... 178

## TABLE OF TABLES

TABLE 1: MOOD PREDICTIONS FROM ATMOSPHERE HYPOTHESIS. ....	31
TABLE 2: THE GERGONNE RELATIONS. ....	40
TABLE 3: PREDICATE CALCULUS EQUIVALENTS FOR PREMISES. ....	50
TABLE 4: NUMBERS OF STATES IN THE OFFICE INDICATOR SYSTEM. ....	64
TABLE 5: FEATURES OF OFFICE INDICATOR SYSTEM. ....	65
TABLE 6: THE VENN COUNTING ARGUMENT. ....	68
TABLE 7: EXPRESSION IN VENN WITH JUST TWO PREMISES. ....	76
TABLE 8: POSSIBLE CHARACTERISTIC DIAGRAMS IN EULER. ....	78
TABLE 9: POSSIBLE DIAGRAMS WITH EULER SOFTWARE IN THE AB-BC FIGURE.....	84
TABLE 10: POSSIBLE DIAGRAMS WITH EULER SOFTWARE IN THE BA-CB FIGURE.....	84
TABLE 11: POSSIBLE DIAGRAMS WITH EULER SOFTWARE IN THE AB-CB FIGURE.....	85
TABLE 12: POSSIBLE DIAGRAMS WITH EULER SOFTWARE IN THE BA-BC FIGURE.....	85
TABLE 13: TWO INFORMATION ENFORCEMENT COUNTS FOR 3 SYSTEMS.....	90
TABLE 14: PILOT STUDY QUESTIONS. ....	103
TABLE 15 TEST ITEMS USED IN SPECIFICITY EVALUATION.....	115
TABLE 16: RESULTS OF SPECIFICITY EVALUATION.....	117
TABLE 17: DIAGRAMS DRAWN BY MATHEMATICS TRAINED VENN USERS.....	119
TABLE 18: EXPECTED THIRD TERM ERROR ANNOTATIONS. ....	121
TABLE 19: EXPECTED STAR MOVEMENT ERROR ANNOTATIONS.....	122
TABLE 20: EXPECTED EMPTY SET ERROR ANNOTATIONS. ....	122
TABLE 21: EXPECTED PREMISE ORDER ERROR ANNOTATIONS.....	123
TABLE 22: VENN ERROR SPREAD BETWEEN LEARNER GROUPS. ....	129
TABLE 23: EXAMPLE DATA FOR EULER MATHEMATICS. ....	131
TABLE 24: THE ERROR SPREAD BETWEEN LEARNER GROUPS USING THE EULER SYSTEM.....	137
TABLE 25: RESULTS OF SPY-RING TEST.....	150
TABLE 26: CARDS USED IN THE CARD-SORT.....	153
TABLE 27: RESULTS FROM CARD-SORTING TASK.....	157
TABLE 28: CARD SORTING RESULTS FOR SINGLE SUBJECT IN INTERVIEW. ....	160
TABLE 29: THE EIGHT CONDITIONS IN ESDA STUDY.....	171
TABLE 30: ERRORS IN 8 CONDITION STUDY.....	177

## Acknowledgments

My principal debt is to my supervisor, Tim O'Shea for the opportunities, help and support he has so freely given during the last six years. Keith Stenning and Bernard Scott have also generously spent much time helping me to clarify my ideas. I am grateful to Diana Laurillard, David Crowe, Norman Udey, Bill Hunter, John Meuller and Tom Carey for helpful comments and discussion at various stages of the research.

The early experiments were conducted with the help and cooperation of the Headteacher, the head of mathematics (Steve Chappel), and other staff and students of Stantonbury Campus school, Milton Keynes, UK. Later experiments with the software were conducted with the active help of Stantonbury Campus school staff, together with graduate researchers and members of the Institute of Educational Technology (in particular Marc Treglown and Janet McCracken).

I was financially supported by an Open University research award while at The Open University, as a visiting scientist at The Alberta Research Council, and as a faculty fellow in The faculty of Education at The University of Calgary.



RESEARCH DEGREES CENTRE  
LIBRARY AUTHORISATION FORM

Please return this form to the The Research Degrees Centre with the two bound copies of your thesis to be deposited with the University Library.

All students should complete Part 1. Part 2 only applies to PhD students.

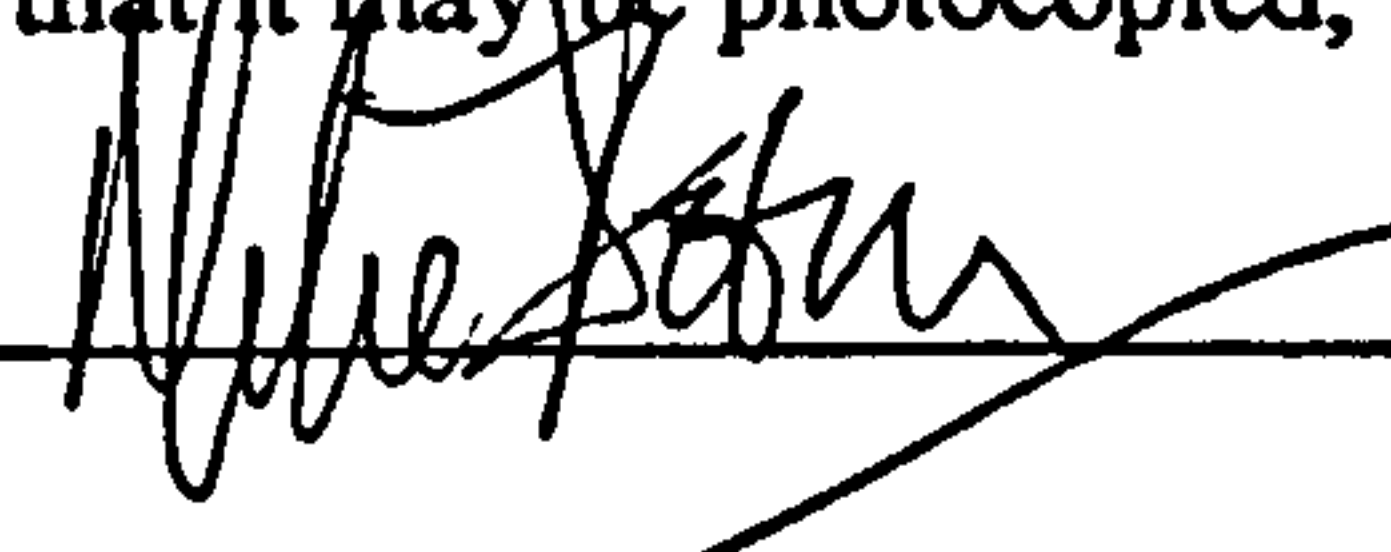
Student: MICHAEL WILLIAM DOBSON PI: 47088257

Degree: PhD

Thesis title: Information enrichment in learning with graphics:  
Improving syllogistic reasoning skills.

**Part 1 Open University Library Authorisation [to be completed by all students]**

I confirm that I am willing for my thesis to be made available to readers by the Open University Library, and that it may be photocopied, subject to the discretion of the Librarian.

Signed:  Date: 1st Jun 2000

**Part 2 British Library Authorisation [to be completed by PhD students only]**

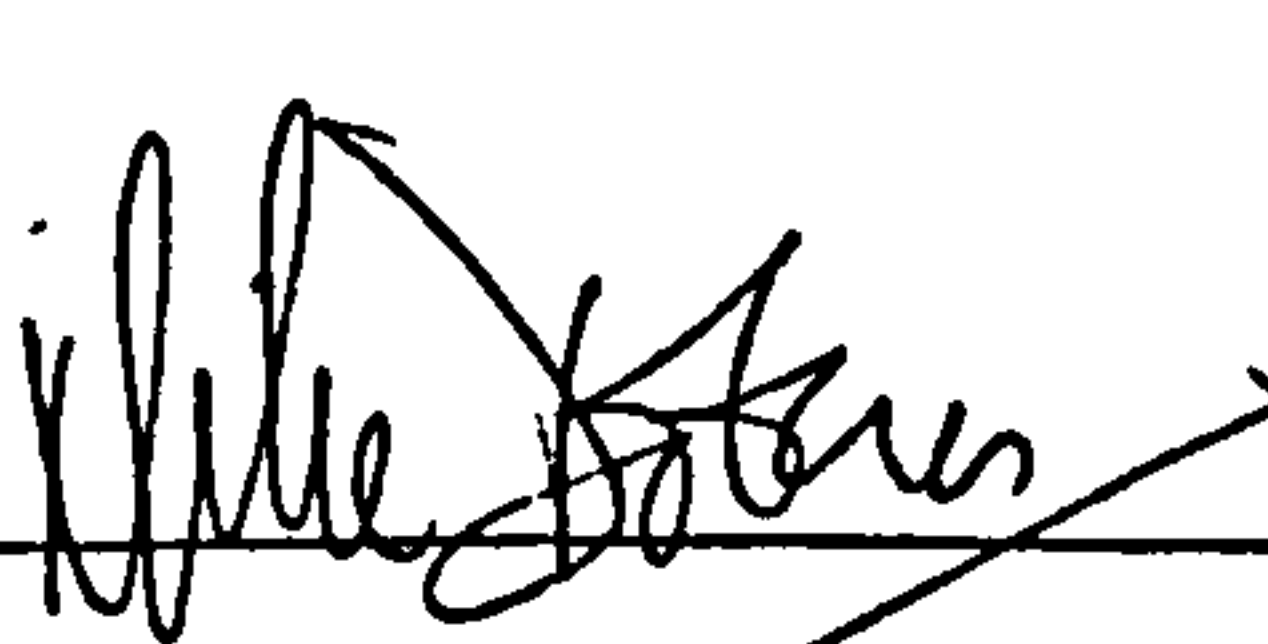
If you want a copy of your PhD thesis to be available on loan to the British Library Thesis Service as and when it is requested, you must sign a British Library Doctoral Thesis Agreement Form. Please return it to the Research Degrees Centre with this form. The British Library will publicise the details of your thesis and may request a copy on loan from the University Library. Information on the presentation of the thesis is given in the Agreement Form.

Please note the British Library have requested that theses should be printed on one side only to enable them to produce a clear microfilm. The Open University Library sends the fully bound copy of theses to the British Library.

The University has agreed that your participation in the British Library Thesis Service should be voluntary. Please tick either (a) or (b) to indicate your intentions.

[a] ☒ I am willing for the Open University to loan the British Library a copy of my thesis.  
A signed Agreement Form is attached.

[b] ☐ I do not wish the Open University to loan the British Library a copy of my thesis.

Signed:  Date: 1st June 2000

## Abstract

This thesis is an investigation into the factors that contribute to good choices among graphical systems used in teaching, and the feasibility of implementing teaching software that uses this knowledge.

The thesis describes a mathematical metric derived from a cognitive theory of human diagram processing. The theory characterises differences among representations by their ability to express information. The theory provides the factors and relationships needed to build the metric. It says that good representations are easily processed because they are more vivid, more tractable and less expressive, than poor representations.

The metric is applied to abstract systems for teaching and learning syllogistic reasoning, TARSKI'S WORLD, EULER CIRCLES, VENN DIAGRAMS and CARROLL'S GAME OF LOGIC. A rank ordering reflects the value of each system predicted by the theory and the metric. The theory, the metric and the systems are then tested in empirical studies. Five studies involving sixty-eight learners, examined the benefit of software based on these abstract systems.

Studies showed the theory correctly predicted learners' success with the circle systems and poorer performance with TARSKI'S WORLD. The metric showed small but clear differences in expressivity between the circle systems. Differences between results of the learners using the circle systems contradicted the predictions of the metric.

Learners with mathematical training were better equipped and more successful at learning syllogistic reasoning with the systems. Performance of learners without mathematical training declined after using the software systems. Diagrams drawn by learners together with video footage collected during problem solving, led to a catalogue of errors, misconceptions and some helpful strategies for learning from graphical systems.

A cognitive style test investigated the poor performance of non-mathematically trained learners. Learners with mathematics training showed serialist and versatile learning styles while learners without this training showed a holist learning style. This is consistent with the hypothesis that non-mathematically trained learners emphasise the use of semantic cues during learning and problem solving.

A card-sorting task investigated learners' preferences for parts of the graphical lexicon used in the diagram systems. Preferences for the EULER lexicon increased difficulty in explaining the system's poor results in earlier studies. Video footage of learners using the systems in the final study illustrated useful learning strategies and improved performance with EULER while individual instruction was available.

Further work describes a preliminary design for an adaptive syllogism tutor and other related work.



# 1 Introduction & Overview

## *1.1 Modality assignment in teaching with graphics.*

The use of graphical illustrations in almost all disciplines of teaching and learning is considered natural and felicitous. Teaching virtually any topic benefits from the regular use of diagrams to elaborate, demonstrate and explain important and often complex concepts. Phase diagrams in material science are essential tools for understanding the heating and cooling of combined substances. Time-zone diagrams help learners grasp the change and passage of time while traveling over long distances. Without logic gate diagrams there would be no way to demonstrate the function of the simplest electronic component. These diagrams are integral to the normal teaching and learning of the topics to which they belong. These diagrams are so natural it may seem there could be no rules to derive from their use except that each must be specially created for its purpose. This work continues on the basis that expertise, however hidden, can be derived about the use of diagrams in teaching.

Choosing the most usable diagram for learning requires expertise that is concealed from the learner and rarely considered by the instructor. The knowledge involved in such decisions may have evolved over long periods of time and may be closely related to representations used in professional practice. For example, the periodic table is a representational system with valuable pedagogic features, but it is also used in schools partly because of its currency with practicing chemists. The instructor sees the benefit of the representations he uses in their ability to vividly communicate the main aspects of the topics they illustrate. Providing explanation for the choice of representation in a teaching situation may seem odd, however this information is required so that teaching systems can be built that will benefit from this expertise. This dissertation is based on a question, which if effectively answered, will contribute to the improvement of computer based teaching and learning.

“What knowledge and expertise can the cognitive sciences reveal about the use of diagrams in teaching and learning, and how might that knowledge be used to design better computer based learning environments?”

One way to characterize the goal of the work here is in terms of “knowledge media” (Daniel, 1996). The idea of knowledge media is summarized as the combination of the best aspects of electronic communications, educationally effective multimedia and the scaleable aspects of artificial intelligence and knowledge representation. Knowledge media will draw on

expertise from the processes of teaching and learning, and a designer may choose to integrate that expertise within the communications, media or adaptive aspects of a design. The expertise involved in choosing the best representation for a learner may be integrated in the choice of media or could be represented explicitly in the adaptive part of a system. If the expertise is stored for adaptive behaviour, decisions about the presentation of material can be made in real time and a system may be able to deal with several topics and different audiences. If expertise is integrated in the choice of media design, it will improve the media for a small audience and will have smaller topic coverage. Whether the expertise is explicitly or implicitly represented, it must be understood so that it can be formalized as rules or guidelines that can later be used by a designer or by the system. Instructors seldom have need for the expertise that is required by a teaching system, because they hardly ever have to make decisions between different representational systems. What initially appears to be a straight forward problem of knowledge elicitation is further complicated by several other factors.

The language used to describe representational systems, whether graphical or not, has been created in many different disciplines and for different reasons. The rules that will eventually describe choices between alternatives must use language that is shared and understood similarly in each rule. Concepts used to talk about representational systems have fuzzy boundaries often without clear distinction between apparently different terms. For example, the idea of “analogical representation” was developed to contrast with representations that are digital. The dial of a compass presents information (bearing to magnetic north) with a representation (the direction of the magnet) proportional with what is represented (the direction of the magnetic north pole). Other terms have been used which for some purposes are very similar. “Homomorphic representations” can for example, be contrasted with numerical (digital) representations, but the author of this term intended it as a set of criteria for systems that use graphical parts, and that are intended to be formally sound. The interdisciplinarity of the research problem requires that a survey of several fields be presented.

There are many disciplines that contribute to the knowledge and professional practice of teaching and learning with diagrams. Among these disciplines are instructional design, cognitive science, learning and instruction, information design and human computer interaction.

The field of instructional design provides guidance in media selection (Romizowski, 1988; Bates, 1995). This pragmatic field has supplied training, particularly in North America, to professional trainers and designers of learning materials for over forty years. Media selection for this group means choosing between print, broadcast television, radio and other forms of storage and review technologies. Technology developments have in many ways, outgrown this kind of media selection support. Since traditional media are becoming increasingly digital, this convergence to one storage and review technology means that



selection questions require a different level of analysis. In spite of the convergence of these traditional technologies, media selection research has remained at a level of granularity that does not provide enough guidance in representational selection.

Cognitive science has produced models to explain why some kinds of representations are more useful to learn about a topic, or to help in problem solving (Paivio, 1971; 1986, Larkin & Simon, 1987). This work has been valuable in demonstrating successful learners' use of graphical systems, and to some extent why useful examples work. These efforts however, are infrequently motivated by the goal of improving teaching and learning through better representation choice. Results therefore, have to be interpreted to generate guidelines for developing computer based educational technology.

The field of learning and instruction continues to add to understanding by contributing new case studies from particular disciplines. The use of graphical systems in learning about geographical time-zones (Schnotz, Picard & Hron, 1993; Lindström, 1980) has led to understanding under what conditions time-lines are used. Sometimes these devices form a part of an algorithm for solving a set problem about time. At other times learners are more likely to use concrete or content related information. In most cases these studies indicate that use of good illustrations is of benefit to the learner. Results from these studies describe the problems students have in learning material and often deal with a mix of representational issues. Studies of this type seldom examine why poor illustrations would have been unsuccessful or what factors make illustrations of benefit to the learner.

In human computer interaction, interface development environments constrain the programmer to building applications that follow simple rules leading to usability (O'Malley, 1990; Norman, 1988). The rules involved in creating 'usable' interfaces are generally based on minimizing additional cognitive load and easing navigation for the user. This knowledge is clearly relevant to choosing illustrations with similar qualities for learners, but translating those rules to alternative options is not possible without an understanding of diagrams.

The question of this dissertation has not been asked before in an explicit way. There are however, many disciplines of research which have contributed to the issue in various ways. The interdisciplinarity of possible sources is an obstacle in itself, because the languages used in each discipline often obscure different meanings and goals, beneath superficially similar questions. Answering such a question within the pragmatic discipline of educational technology is a fairly typical problem within the field because the practical nature of most problems within educational technology must also draw on many fields of knowledge. The measures of success in this field are determined by the development of the technology for the learner and the improvement in learning that is enabled. The nature of the field also prescribes an eclectic and exploratory research approach. The methodology used in this work looks for answers to specific questions in controlled settings and uses rigorous analyses to

support the research question and to interpret results. The method also allows for unexpected results to furnish a deeper understanding of the topic and how learners deal with it.

Research methods involving the measurement of learning necessarily require a topic to learn even when the research objective is generic and intended to be applicable to several topic areas. For reasons which are outlined below, 'syllogistic reasoning' was selected as the topic area for this project. Although the work provides valuable results that help to directly answer the research question, there are two other sets of results that are equally valuable. The first of these describe the regular and predictable errors that learners demonstrate while using the diagram systems to learn syllogistic reasoning. The second set of peripheral results concern differences in abilities between learners with training in either mathematics or humanities topic areas.

The following questions provide a framework for the rest of the dissertation.

1. How can multimedia teaching materials be precisely described, so that a teaching system can decide on optimal media for a learner?
2. How can cognitive approaches to learning with multiple representations, guide system design in deciding between alternative media?
3. Which research methods from fields contributing to educational technology can serve its goals with practical and reliable benefit, and how can these methods be shaped into a coherent methodology for the discipline?
4. Which variables, in the design of interactive graphical tools, directly affect learning and what are the optimal values and weights of those variables.
5. What difficulties and misconceptions do learners encounter while learning syllogistic reasoning with and without the graphical notations and the computer based support tools?
6. Do cognitive approaches help to explain learners' success with different notations, different software tools and different learner behaviour?
7. What has been learned coincidentally, what are the limitations of the results, and what future work will form the basis of continuing research from this thesis?

In the remainder of this chapter, the thesis is reviewed chapter by chapter illustrating the chronological progress of the project, the development of goals, the evolution of study plans and finally, suggestions for continuing future work.

Chapter 2 reviews current understanding in cognitive and learning sciences about the uses and benefits of graphical representations in learning. The chapter surveys work from several programs that have a bearing on the questions in Chapter 1, and provides a framework for empirical study and interpretation of results.

Qualitative representations are simplified versions of accepted knowledge that are sometimes more like naive learners' models. These representations are often presented using informal graphical languages, but sometimes with formal non-graphical languages. Graphical systems used in learning are mostly informal although certain strict criteria make their syntax rigorous. Some graphical systems, called 'homomorphic,' are formal and pay special



attention to the similarity of form between the referent and the graphical lexicon. This property of representations enables situated environments. Learning from these graphical systems involves building a manipulable structure that can be modified and added to in a way that will predict new outcomes. Learning inevitably includes progressing from poorly constructed and even contradictory models towards models that are computable and identical to the structure of the domain. Multiple models may be necessary for real world reasoning. Understood correctly, the diagram provides a set of constraints on which the learner can continually test their model progression.

The benefits of diagrams are explained by “computational”, “architectural”, and “special purpose” accounts. The computational account suggests measurable characteristics of diagrams. This account provides the best opportunity for creating rules usable by designers or by automated computer based learning environments.

Chapter 3 is the first in which syllogistic reasoning is elaborated and discussed in detail. Until Chapter 3, the discussion of learning with diagrams has been free of any commitment to a particular topic area. In this chapter a decision is made to use this topic in the studies to follow. This decision was made for several reasons.

Depending on how exactly they are counted, there are in the order of 64 forms of syllogistic problems. The size of the domain is constrained enough for study designs to be well controlled, and large enough to avoid learning by recall or simple recognition. Other researchers particularly in cognitive sciences have thought similarly about the domain and for this reason there is a long history of analysis in human performance of syllogistic reasoning. This data enabled many cross-checks with the studies here, and even led to the design for a modification of a method of adaptive testing outlined in the final chapter. Most of the previous work in syllogistic reasoning, focuses on explanation for people’s failure to perform the skill. The overview in Chapter 3 shows that these reasons are likely to be countered by the use of graphical systems. Most of the reasons fall into one of two kinds. The first kind of reason is confusion about the meanings of words that are used but have a technical sense in reasoning or logic. The second reason is based on the difficulty of remembering steps in processes or algorithms used to reach conclusions. The graphical systems may help to reduce the numbers of those steps, and provide a simplified structure for learners to use in their heads.

The chapter describes seven systems that can be used to demonstrate and support learning syllogistic reasoning: (a) Venn Diagrams, (b) Euler circles, (c) Edinburgh Euler, (d) Johnson-Laird’s mental model notation, (e) the predicate calculus, (f) TARSKI’S WORLD, and (g) Lewis Carroll’s ‘Game of Logic’. These systems work in a variety of ways combining circular enclosure, intersection, labeling, and a combination of spatial layout and annotation.

The last part of Chapter 3 identifies key differences between the systems that are likely to affect learning. A model or principle of information processing is introduced called the

specificity principle. This principle describes several factors that are different between systems. Factors which are not covered by the principle can be controlled for in study or treated separately. Although the principle of specificity appears to cover many of those factors, it is already clear that an open-ended, mixed and opportunistic methodology is most likely to bring results.

Chapter 4 begins with an explanation of the specificity model. A simple and readable example illustrates the main constituents of the model. The example uses a situation in which a shared office may hold several workers and a visitor wants information about those in the room without entering it. Four scenarios develop the model of specificity by illustrating the differences between them. The models describe two related factors in a representational system, its expressivity and its specificity. The breadth of information that a system can illustrate counts as its expressivity value. The specificity scale is determined by comparing the breadth of information communicated with the upper level of possible expression for that system. The specificity model suggests that systems with unwieldy or cumbersome expressive abilities will have less value for the learner than those which are more commensurate with the scale of the domain. The chapter analyses four systems that have fixed upper limits of representational states: Venn, Euler, Edinburgh Euler and Carroll. A simple wire-frame cube labeled at each vertex and edge, enables counting the number of possible states in the Venn and Carroll systems. The discovery of the relationship between the cube and the Venn diagram is a valuable finding. Understanding the reason that the cube is useful in the counting process also adds to our understanding of the value of 'intermediate representations' more generally.

Chapter 4 ends by using results of the specificity counting process to predict which of the systems is most likely to be useful for learners. The Euler system scores lowest of all the systems and is therefore the most specific and should be the most useful. Euler is closely followed by Venn and the Carroll systems. The remaining systems: (a) TARSKI'S WORLD, (b) Johnson-Laird's notation, and (c) the predicate calculus cannot be counted using the specificity model. These systems have an infinite total of possible configurations.

Chapter 5 is the first of three chapters that present plans and results from empirical studies. This first study helps to establish the possibility for more extensive data collection in the following studies. Three software applications can be used to support syllogistic reasoning: (a) the TARSKI'S WORLD (Barwise & Etchemendy, 1991), (b) VENN (Venn, 1991) and (c) EULER (Stenning & Inder, 1995; Stenning & Oberlander, 1995; Stenning & Tobin, 1993). The description of each application explains its main features and operations. This description focuses on the software and not the notations. Chapter 3 describes the abstract systems and algorithms used to solve problems.

There are conspicuous differences between these applications that would affect the comparability of experimental conditions using them. Some software offers long lists of



examples and remediates incorrect responses from learners. Without offsetting these differences, experimental study would be more likely to assess the merits of each application rather than the strengths of the graphical system that each application uses.

The pilot study used three subjects from a local secondary school. The study took place in a teaching room on the campus of the Open University. The learners and the instructor engaged in an open-ended instructional discussion about syllogistic reasoning. The instructor presented the language used in the topic area with careful definitions of each important word. Example problems were presented on a white-board using each of the graphical systems. The group discussed each system, exploring component parts and differences with other systems. Finally, subjects were able to explore the topic with the interactive software. During the three hours, interactions between the instructor and the three subjects were recorded on audio tape. Transcriptions of these tapes illustrate examples of misconceptions and errors in the dialogue.

The study revealed that these particular students could engage in the topic area and were able to demonstrate competence towards the end of the interaction. The transcriptions identified learner difficulties that were consistent with reasoning 'biases' from the reasoning literature. Each learner was strongly motivated to engage with the software and showed continuing interest over more than an hour. There was some evidence from this pilot that the TARSKI'S WORLD application proved difficult for learners to use. Interest in reasons for this difficulty persuaded us to include the application in the following round of studies.

The pilot study supplied several 'lessons learned' that were incorporated into the following studies. The transcribed instructional sequences led to five instructional booklets, one for each of the systems and one general booklet describing the language of syllogistic reasoning. These booklets: *An Introduction to Aristotelian Logic*, *The Euler system and the syllogism*, and the others, all contain in-text activities and questions that replicate the more successful aspects of the dialogue from the pilot study. Off-setting the functional differences between each of the applications involved creating a list of syllogistic problems for learners to use as examples.

Chapter 6 describes a study in which 42 learners used three different software applications to learn syllogistic reasoning. Pre to posttest scores on each of the systems provided a measure of each systems' benefit to the learner. Since each of the systems had already been evaluated with the metric, the results were easily used to assess the value of the metric. Results supported the specificity principle while comparing systems that were radically different in their ability to express information. The most expressive system of the three, TARSKI'S WORLD, was the least useful to all learners in the study. Comparisons between the two less expressive systems (EULER & VENN), were however much closer and did not support the specificity principle. The two remaining findings were incidental to the objective of the study, however they are equally of value.

Subjects were allocated to one of three conditions based on the three alternative software applications. The samples in each condition had equal numbers of learners with interests in different topic areas. One group of subjects were taking advanced mathematics level studies and mostly science subjects, and the second group were taking mostly geography, history and English classes. After the study, these conditions were separated and a large difference was found between the effects of using the software in each population. Most of the 'humanities' subject scores declined from pre to posttest. By comparison, those learners with mathematical backgrounds benefited substantially from their interactions with the support tools.

The second, peripheral set of results provides detailed data that may be directly used in design of new systems for learning syllogistic reasoning. The posttest questions from this study, were identical in form to the pretest questions in every way except one. In the posttest, subjects were also asked to construct the diagram that best represented the problem. From this data it was possible to interpret a number of detailed errors shown by learners. Each error was given a name and the analysis generates a table, showing the exact errors that would be expected for each type of problem in each of the graphical systems. These results will be useful in the diagnostic part of a tutoring system to teach syllogistic reasoning. These detailed tables of expected annotations point to the possibility of a generative model of diagnosis for this topic area. The future work section in Chapter 8 describes part of a plan to continue this work.

Chapter 7 explores two possible explanations for results of the studies reported in earlier chapters. The study was designed to explain differences between mathematics and humanities students, and the failure of the metric to predict learning outcomes between systems with similar expressivity. Four kinds of data were collected with 20 subjects.

A card sorting task helps assess learners' difficulties in understanding non-dynamic component elements of representations. These differences may impact on the outcome of the specificity metric. Learners' prior conceptions of graphical elements are independent of any computational measure of the representations. These conceptions may therefore constitute the confounding variable that helps explain the failure of the metric to predict between systems of similar expressivity.

The card sorting task uses representations of single premises required for each syllogism and so does not require understanding of the inference mechanism for each system. The card sorting task is similar to eliciting student views through a technique called "interview about instances" (Gilbert, Watts & Osborne, 1985). The instrument used is a modification of Gilbert's approach and provides the additional benefit of ranking learners' preferences for individual components. The study shows that representational conventions vary in their ease of understanding because of learners' preconceptions before any training has taken place. The influence of expectations, however, did not help to understand why the



specificity metric failed to predict earlier results. Learners demonstrated a notable preference for conventions from the most specific system. It was therefore not possible to isolate 'prior preferences' as the missing confounding variable from the previous studies.

A second study used the spy ring test (Pask & Scott, 1975) and examined differences in approach to learning by learners with training in either humanities or mathematics topics. The study sought to explain the earlier findings that mathematics students improved in their reasoning skills after using the software systems and that humanities students' performance declined with the same treatment. The spy ring test discriminates between 'operationalist' and 'comprehension' type learner styles and originated in Pask's early work with the CASTE system (Pask & Scott, 1975). The study shows that learners with mathematics training show a strong correlation with a serialist approach to learning.

A third study makes use of exploratory sequential data analysis or ESDA (Harrison, 1995; Sanderson, James, & Seidler, 1989) with video of students solving syllogisms with and without the software. Through analysis of the time based behaviors using TIMELINES software, the study elaborates on time related aspects of known errors and demonstrates learners' progress during instruction. Together with work-scratchings collected from this study, this data set helped to categorize errors and to build a concept map that gives a summary of learner difficulties with the topic.

The final chapter reviews the contributions of the thesis and the limitations of the results. The results of the thesis are relevant to several groups. Cognitive scientists may find the metric useful for their own models and to apply empirical assessment as a criteria to test their theories. Logic teachers and developers of systems for teaching logic will draw from the catalogue of bugs shown by students with different backgrounds and abilities. The same group will be interested in the review of software that can be used in teaching logic, and the treatment of designs that have a positive affect on learning. Curriculum planners may note that a mathematical training prepares learners with skills that lead to problem solving whereas other disciplined may not. It may be fruitful for educational technology researchers to develop and use the card sorting method presented in Chapter 7. This final chapter shows that the eclectic and practical approach developed in this thesis may be a model for others to adopt.

Chapter 8 also describes a range of future work projects. One project arose during the instructional phase of the pilot study. Reflecting on the experience helped to see that problems were presented to learners in order of increasing difficulty. Progression of this kind would suggest the use of adaptive testing algorithms. Although previous studies in syllogistic reasoning collected difficulty data for each problem type, their analysis provides data that is unusable for normal adaptive testing processes. Chapter 8 explores a modification of an adaptive testing algorithm called sequential probability ratio testing (or SPRT) that can accommodate data in this other form. The use of these adaptive presentation techniques in



teaching the topic area will enhance the quality of interaction and the motivation of learners using them (Huang, Collins, Greer & Dobson, 1995). Other projects outlined in Chapter 8 include, automating more of the counting process, developing teaching systems with multiple representations, improving the metric, and training knowledge workers with formal reasoning skills.

Chapter 2 elaborates on current understanding of diagram use in learning, and provides a basis for the program of analysis and study that follows.

## 2 Learning with Diagrams

Progress in the design and development of educational technology must draw from many different fields of knowledge. Improving learners' experiences with computer based graphical representations demands investigation in several areas. The goal of this project is to precisely describe the features of graphical representations that benefit learners. This information will better inform the design of diagrams in courseware. In the longer term the courseware will use this information to make decisions about representations for the learner.

Instructional design is a research and training discipline that has approached the related issue of media selection. Recently, Dijkstra (1997) asserted: "There is no best medium in education, but generally it is quite plausible to assume that for the acquisition of particular knowledge and skills, some media will be more adequate than others." Dijkstra, Seel, Schott, and Tennyson (1997) suggest the need for a process oriented media selection model and maintain that the model should extend to more fine-grained support of micro-level decisions about multimedia elements. Previous work in media selection for distance education (for example, Laurillard, 1993, Bernsen & Bertels, 1994; Bates, 1995), has effectively described some of the important attributes of alternative technologies used for delivery. There are plainly intersections in the communication possible with these technologies (broadcast radio, VCR tapes, television programming, *etc.*). Diagrams, for example, may appear in any of the visual media. As these technologies converge towards a digital standard, the need for decision support at a finer granularity becomes increasingly clear.

The cognitive disciplines hold a range of views about the processes involved and the advantages of using diagrams in learning. There is broad recognition that an effort to better understand the processing of information represented with visual representations is an important topic of study (*e.g.*, Chandrasekeran, Narayanan & Iwasaki, 1993). Work to achieve deeper insight into processing of diagrams has taken place both in analytical and empirical frameworks. Development of understanding in any area can progress only when there is a clearly agreed language with which to discuss, hypothesize and test claims made in experimental situations. For communication within intelligent tutoring systems the language framework must be clarified within a general ontology that enables action as well as understanding (*e.g.*, Gruber, 1993; Bourdeau & Mizoguchi, 1999). The pursuit of an analytical framework for graphical systems is especially important when the goals of research involve using the framework in computer based decision making. A designer using guidelines developed from an analytic framework may be more tolerant of ambiguities in the language.

For a computer based tutoring system to make effective choices between representations of the material it must begin with a commonly understood language describing the field.

Visual representations are used widely in many different areas of human activity including teaching and learning. Diagrams have often proved to be effective in the process of invention and creative thinking. Since we are all users of diagrams both in work and in recreation, it is not hard to understand their appeal. Proponents of diagram use often assign them with spectacular qualities. Charles Pierce, a late nineteenth century logician and inventor considered *all* valid reasoning to be diagrammatic by definition.

Mathematical reasoning consists in constructing a diagram according to a general precept, in observing certain relations between parts of that diagram not explicitly required by the precept, showing that these relations will hold for all such diagrams, and in formulating this conclusion in general terms. All valid necessary reasoning is in fact thus diagrammatic (Pierce, 1903).

Part of the reason such a claim is possible is that pinning down a necessary and sufficient definition for graphical systems is so problematic. The first set of questions in chapter one ask how to realize precise description of multimedia elements, especially graphical representations. This chapter discusses research in visual representation from the cognitive and learning sciences. The process begins by discussing the definition and limits of graphical systems.

### *2.1 Defining graphical representations.*

Classic examples of graphical representations like the phase diagram, the periodic table and the Cartesian graph, all help to exemplify the kind of representation at issue. These examples however, rely on several graphical conventions that differ from example to example. Probably the most central feature of graphical representations that distinguishes them from others, is the use of spatial cues to display and communicate information. Again, the use of these cues varies from one representation to another. A sketch used to plan the layout of a kitchen clearly uses spatial cues that are very similar in form to the way the kitchen will appear if the plan is followed. By contrast, the use of spatial reference in a mind-map allows similar concepts to be placed near to each other and sometimes linked by a line or an arc. Relations between objects shown in the mind map are much more abstract than in the kitchen plan. Unlike the finished kitchen, the idea of semantic relationships between concepts is abstract and intangible.

Using spatial layout as a criteria to define graphical representations causes difficulties for distinguishing between graphical and non-graphical representations. In truth, most representations are a combination of graphical and non-graphical components. Even those representations without any obvious textual component, must be explained in the context of a



natural language. Diagrams are usually accompanied by labels and annotations, and text is often arranged with spatial cues such as indentations, variable font sizes and bulleting. Opponents of diagram use for specific purposes are sometimes portrayed to believe that diagrams are distinguishable from other kinds of representations by their inability to represent information in a formally sound, unambiguous and valid way (Barwise & Etchemendy, 1996). The discussion below shows this belief is not however true. Graphical representation will be treated for the remainder of the project as a concept under development. The project helps to add to a definition for graphical representations by reviewing several approaches from different disciplines.

Diagrams are often used in teaching and learning to communicate a simplified version of a physical system. The representation may only be adequate to illustrate a small part of the system, but for this reason, can be an effective beginning for a naive learner. These qualitative diagrams are an effective instructional strategy for teachers and learners of scientific material.

### *2.1.1 Qualitative diagrams in science.*

In science teaching, speculation and study often concern the accuracy and completeness of diagrams. Representations used in science teaching are often described as either qualitative or quantitative. These terms are not interchangeable with graphical and non-graphical. Nevertheless, diagrams of physical systems frequently illustrate qualitative material and rarely illustrate quantitative information.

Most scientific subject matter presents more scope than logic for developing instructional strategies based on detail and accuracy of models presented to the learner. In complex scientific domains, for example, cardiovascular pressure systems, tutors and students in normal teaching and learning environments will switch between levels of qualitative explanation (Khuwaja, Evens, Rovick, Michael & Patel, 1996). A teaching strategy based on this observation could lead to a set of general guidelines for using representations in learning. Qualitative representations are closer to the common-sense reasoning of naive learners than the precise and detailed models that correspond to accepted knowledge in the field. Recent work (De Jong, Ainsworth, Dobson, Van der Hulst, Levonen, Reimann, Sime, Someren, Spada & Swaak, 1998), suggests ordering effects in representational choice. Learners working in physical sciences tend to benefit from sequences where qualitative representations come before formal descriptions. Qualitative representations benefit the learning process by providing comprehensible versions of topic material that modulate accuracy and depth.

Qualitative representations are by no means always graphical. Nevertheless predominantly physical systems are often represented with diagrams, and often with limited and elementary coverage of a system. An example domain is provided by structural analysis.

Accurately representing the application of a large force to a load bearing member in a building involves calculating a complex set of factors. Using this model to predict contortions and structural failures would require precise measurements of beam thickness, material strengths and loading. This example (Tessler, Iwasaki & Law, 1993) demonstrates how the concept of qualitative representation is used somewhat interchangeably with graphical representation. Tessler proposes three important features of qualitative diagrams in structural analysis. Each of these features, with simple modification, applies to a more general model of graphical representations. These representations imply a predictable sequence of behavior from an initial input. Although this example has not been implemented as computer based learning materials, it is strongly suggestive of an interactive graphical representation. The diagram is a visual language of (structural) behaviour that requires a very limited textual commentary for understanding. Graphical representations may present information from any topic area however Tessler's analysis applies best to domains where there is a clear underlying model of interaction between system variables. These diagrams allow the designer to retrace sequences of effects from initial causes. The diagram acts as a short-term memory device that stores previous deformations of the structure as states in the changing diagram. Diagrams in general, can lead the user to see where input will alter system states and provide a simple language with which to continuously update illustration of the current state.

Tessler's analysis reports that the qualitative diagram illustrates a simplified version of the substantive physical system that allows the user to judge the next place to look in solving a problem.

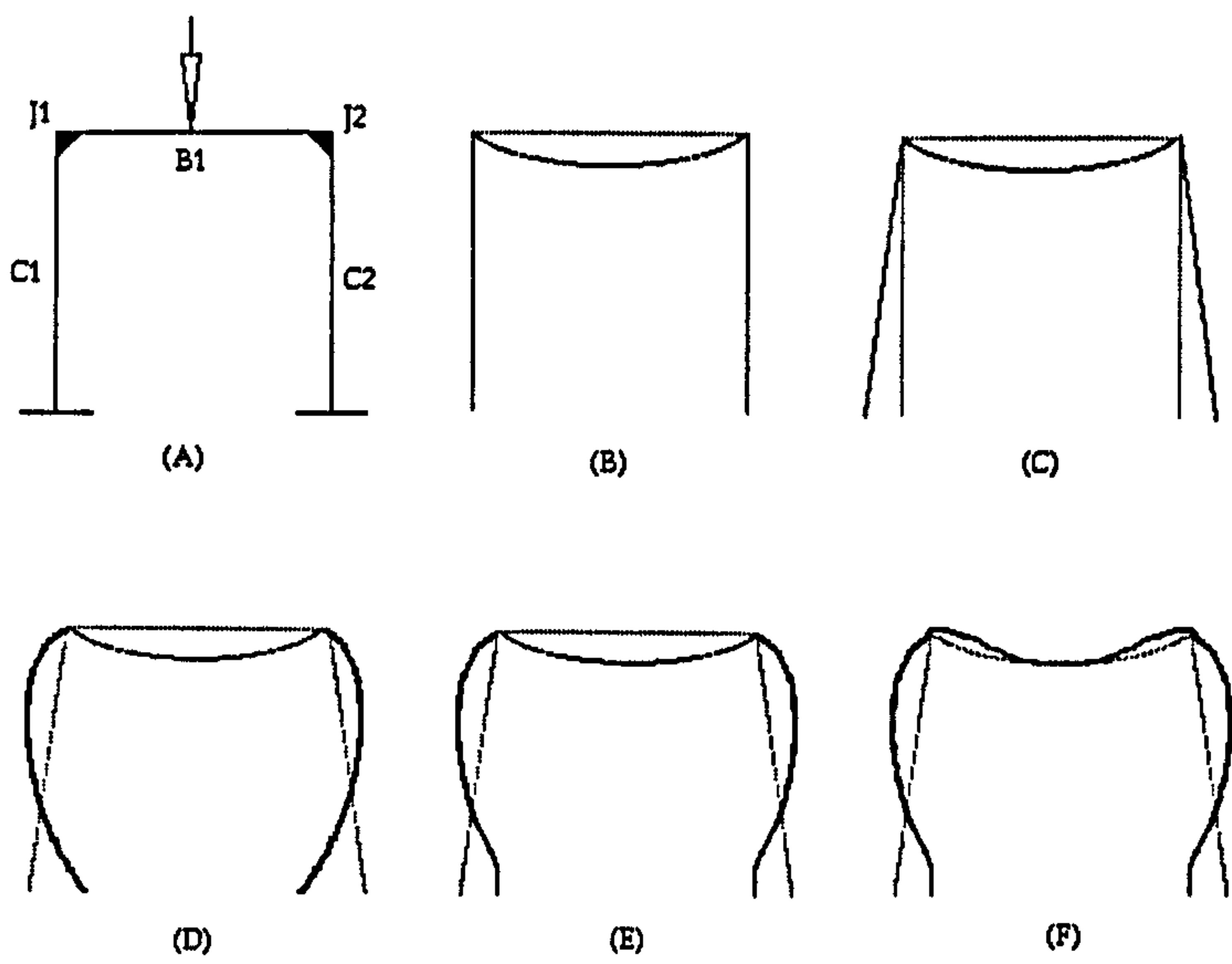


Figure 1: Qualitative stages of thought  
(from Tessler, Iwasaki, & Law, 1995).

A typical sequence of qualitative reasoning from Tessler's account is reproduced in Figure 1. The first picture (A) shows a load being applied to beam J1-J2. This load deflects the beam in the negative 'y' direction which leads to the situation pictured in diagram (B). However columns C1 and C2 must remain perpendicular to J1-J2 because they are rigidly connected. A new diagram (C) is drawn that preserves the 90° angles between the horizontal and vertical beams. This picture does not preserve the fixed points at the foot of columns C1 and C2. Restoring the fixed points of C1 and C2 produces diagram (D). The process continues until all structural constraints are resolved. Diagram (F) represents the final effect of the initial load.

The success of qualitative representations in learning and problem solving may be the result of several factors. These representations present the minimum necessary information for the problem in a form that is close to the naive learners understanding of the topic. The small lexicon involved in these qualitative diagrams corresponds to the minimum necessary features required for making simple inferences about the problem. The lexicon is strictly inadequate to make decisive inferences about any particular situation. To understand the process in an abstract way, and for generating alternative possible outcomes for a real situation, diagrams provide adequate guidance (Larkin & Simon, 1987). The diagram does not always need to take the part of a fully systematic method in itself but can provide an anchor or qualitative picture of the domain's most simple and salient features. This anchor is added to during the problem solving process and used to check intermediate steps.

The structural deformation example uses qualitative representation explicitly to show how these limited models of physical processes can help in problem solving. It would be inaccurate to say that qualitative representations are always informal. Earlier work in qualitative process theory (Forbus, 1983) presents qualitative illustrations accompanied by formal descriptions of the naive physics described by the diagrams. Indeed, one of the goals for qualitative process theory, was to better understand naive models of physical processes by describing them with formal languages.

Qualitative representations help learners to understand a topic through providing limited and strictly inaccurate and imprecise information. Qualitative reasoning and representation is indifferent to the mode of representation but has often used diagrams to illustrate naive views of physical processes. These diagrams and the reasoning they represent, have sometimes been cast in formal descriptions with logical languages. Qualitative representations are therefore also indifferent to the distinction between formal and informal languages. While the languages used to describe naive physics are formal, the diagrams are not. Many diagram systems measure up to the strict criteria of formality but most used in teaching and learning of scientific topics do not.



### 2.1.2 Formal graphical systems

Research in mathematics and logic teaching investigates the use of diagrams in learning. Opponents of diagram use identify the practice with the ‘withering away of formal semantics’ in favor of less rigorous approaches to formal study (Tennant, 1986). Tennant’s major complaint is that graphical representations are inadequate and cannot play the same role as traditional notations. Tennant’s (1986) criticism of diagram use in favor of formal semantics held that diagrams are not capable of the formal precision expected from traditional languages.

Diagrams are a commonly used instructional tool in mathematics, although usually accompanied by textual notations for categorical proofs. It is a newer phenomenon to see diagrams used as rigorous formal systems of proof in their own right. The diagrams used in the empirical studies reported in later chapters each support an exacting semantics as sound and valid as any other. These would not have suffered Tennant’s complaints. Opponents to diagrams in teaching formal systems are *de facto* opponents of informal representations.

Certain diagram systems are just as formal as any textual notation system. Precise meanings for the diagram lexicon and allowable transformations are described in an agreed, usually non-graphical formal notation. These systems satisfy the conditions for any formal language. Venn diagrams, blocks world diagrams, geometry, circuit diagrams, Pierce’s logic and Euler circles have all been described in this way (Chandrasekaran, Narayanan & Iwasaki, 1993; Shin, 1991; Barwise & Etchemendy, 1990; Wang, 1995). According to these writers, the formality of the representation depends on neither the psychological modality (visual, aural, *etc.*) or the representational mode (e.g., text or graphics) of the notation, but simply on the strictness of its syntax and semantics. Graphical systems that are sufficiently defined to meet the criteria are as formal as any other language. These standards are strict and most uses of graphics in teaching and learning do not meet them. The graphical systems used in the empirical studies reported in Chapters 5, 6 and 7 are unusual in that they meet with virtually all of these criteria. The standard for formal systems below uses the example of chess notation and comes from Goel (1992;1995). The standard has six propositions, or criteria:

- *Syntactic Disjointedness*. Each token belongs to at most one symbol type. Thus for example no tokens of type “rook” belong to the type “queen.”
- *Syntactic Differentiation*. It is possible to tell which symbol type a token belongs to.
- *Unambiguity*. Every symbol type has the same referent in each case and every context in which it appears, no bishop refers to a knight regardless of context. (Even as the pawn reaches its end rank and becomes a Queen, it is not both Queen and pawn at the same time).
- *Semantic disjointedness*. The classes of referents are disjoint, that is each object referred to belongs to at most one reference class. For example no pawn belongs to the class of rooks.



- *Semantic differentiation.* It is possible to tell which class a particular object belongs to. Given a king and two classes of objects one could determine which class if any the king belongs to.
- *The rules of transformation of the system are well specified.* There is no question of what does or does not constitute a legal move for a bishop. The legal moves or transformations of the system are such that these properties are preserved at each state.

These standards evidently help to distinguish between graphical systems that are formally rigorous and those which allow ambiguity. Systems that meet the standards can be used as formal languages to describe many different domains including chess, geometry and circuit design.

### 2.1.3 Homomorphic representations.

Two standards for representations focus on the similarity of form between the language used and the objects being represented. From a learning and instructional point of view, such a standard is attractive. The theory of cognitive apprenticeship (Brown, Collins, & Duguid, 1989) for example, states the need for authentic representations of phenomena

Two key terms accurately describe representations that are similar in form to the objects and relations they represent. Barwise and Etchemendy (1995) describe the term *homomorphicity* and Sloman (1993) describes *analogical* representations, Kulpa (1995) compares the two. The study of analogical representation mainly concerns the need for artificially intelligent systems to function with diagram-like mental structures, since natural intelligence uses these tools so extensively. An important goal of artificial intelligence has been to create working models of human intellectual processing that not only fulfill behavioral criteria, but also qualify as plausible descriptions of natural intelligence. The discipline has mainly used symbolic languages that excel at processing lists, performing set-theoretic manipulations and supporting data structures like scripts, schema, and frames (*see* Charniak & McDermott, 1984; Rich, 1988). However, Sloman (1971) has argued that much of human reasoning is not symbolically based and that analogical representations are a large part of human cognition (Sloman, 1971; Sloman, 1993). The importance of analogical representation in human processing is recognized and has driven the renewal of international collaboration and effort (Glasgow, Narayanan & Chandrasekaran, 1995).

Analogical representation is similar to homomorphic representation except that the criteria for homomorphic representations are more precise and formal. Sloman's work emphasizes internal representation both in the natural intellect and in computational models of diagram processing. The emphasis of the Barwise group is on finding external representations that are effective for learning first order logic.

Features of homomorphic representations according to Barwise and Etchemendy (1995) are as follows:

- 1) *Objects are represented by icon tokens. Each object is represented by a unique icon token and distinct tokens represent distinct objects;*
- 2) *There is a mapping  $\Phi$ , from icon types to properties of objects;*
- 3) *The mapping  $\Phi$ , preserves structure, that is;*
  - a) *If an icon is a sub-type of the other (as in the case of a shaded square and square), then there is a corresponding sub-type property relation among properties they represent, and,*
  - b) *If two icon types are incongruous (say, squares and circles), then the properties they represent should be incompatible.*
- 4) *Relations among objects are represented by relations among icon tokens, with same conditions as in 3a and 3b;*
- 5) *Higher order relations among objects (like transitivity, asymmetry, reflexivity are reflected by the same properties of relations among icon tokens;*
- 6) *There are no possible situations that are represented as impossible, and,*
- 7) *Every representation indicates a genuine possibility.*

The homomorphic standard is tied to a belief about learning and problem solving that is widely held. Learning an abstract process or idea, benefits from the use of examples that are authentic and concrete. Constructivism (e.g., Jonnasen, 1991) and situated cognition (Brown, Collins & Duguid, 1989), both emphasise the idea that knowledge is bound to activities in action, and that the environment of activity constitutes the only reality for the learner. Within these frameworks, learning environments for formal logic would meet the criteria for homomorphic systems.

#### *2.1.4 Modeling knowledge with diagrams.*

Developing understanding in a domain requires the learner to form a mental model that can be manipulated in order to answer questions relating to the domain (Reiser, Kimburg, Lovett & Ranney, 1992; Du Boulay, O'Shea & Monk, 1981; Johnson-Laird, 1983; White & Fredericksen, 1990). Models include process as well as semantic information. A complete model of a system must include an understanding of the progression of states that occur between an initial input and a final answer (Johnson-Laird, 1983). Depending on the system being modeled, these states could be changes in data during the execution of a program, or change in electrical flow across components in a circuit. A fully developed model is developed to such complexity that the propagation of the underlying causes of phenomenon and their effects, can be extended to predict the result of added procedures, variables or other model components.

Norman (1983) points out the significance of mental models in the conceptualization users bring to learning new software interfaces. He reminds us that learners have both unstable and incomplete models of any system. Tutors also have models that describe how learner's models compare to their own or with expert's models. Norman concludes that the nature of learners' models has consequences for supporting their development. "People's mental models are apt to be deficient in a number of ways [we must] learn to understand the



messy, sloppy incomplete and indistinct structures that people actually have” (p.14). Gentner and Stevens (1983) study people’s mental models of electricity and conclude, “Although initial models may be fragmentary, inaccurate and even internally inconsistent, nonetheless they strongly affect a person’s construal of new information in the domain” (p. 126). Lewis (1991) sees learning as the process of refining models so that they *correspond more closely with the real world*. According to Lewis these models are arranged in the mind of the learner in some hierarchy corresponding to a semantic network of sorts, and each part of the network is able to be deleted, rearranged, generalized, or specialized as learning and model development improves.

Perhaps the most rigorous treatment of mental models deals with failure and development in human reasoning (Johnson-Laird, 1983; Johnson-Laird & Bara, 1984; Johnson-Laird & Byrne, 1991). This discrete model is concerned with simulating human problem solvers. “We do not seem to use logic to solve problems and yet we are capable of reasoning which surpasses the meager constraints of the most complex of computer programs” (Johnson-Laird & Byrne, 1991). The account of Johnson-Laird and colleagues, proposes constraints on what can constitute a mental model. There are nine constraints in all, but two relate directly to the topic here. The construction and interpretation of the model should be computable, and in its fully developed form, the structures of the model are identical to the structures of the states of affairs that the model represents. These models can lead to valid reasoning, although they also explain the errors that humans make. As Boden (1988) puts it: “whereas formal logic is a culture-specific cognitive resource, (which even professors of logic often ignore), the ability to construct and reason with mental models is a general feature of the human mind” (p. 177). Johnson-Laird’s (1988) account of natural reasoning, shows that valid reasoning is equivalent to the construction and manipulation of the possible worlds that can be created from the interpretation of a sentence. The theory defines procedures for seeking counter examples and for revising previous models. In this way the theory implies a certain kind of interaction with a representation. Johnson-Laird’s is the only mental model approach which is suggestive of an instructional sequence or form of strategy. The process is demonstrated in Chapter 3 where it is compared with several other approaches to understanding human reasoning.

## 2.2 Diagram processing.

Charles Peirce (1903) described the role of diagrams in creative thought, invention and discovery.

*The diagram becomes the something (non-ego) that stands up against our consciousness. In the same way that there is an inherently dyadic quality to perception, so too are we aware, in mathematical reasoning, of the world which forces itself upon*

*us. Creativity is the cooperation between diagrammatic insight and logical thought; without a verbal and logical description our new images and hypotheses would be useless. Genius is the ability to combine both insight and logic. Kekules's discovery of the Benzene ring or Archemedes' discovery of the notion of displacement, began with visual images but would have ended there if it had not been for subsequent logical analysis.*

The cooperation of the diagrammatic, natural language and formal logical modes, all contribute in Peirce's view, to a productive self moderating method of inquiry. Three more recent accounts of diagram processing add more detail to this attractive idea. The first explains the benefit of diagrams in terms of their tendency to be 'special purpose' designed for particular uses. This approach suggests that the benefit of diagrams for the learner is a result of environmental conditions experienced by the human mind. Learners continually act in a spatial world that provides the habituation of reasoning that translates into effective use of diagrams. Because of this familiar reasoning, diagrams begin to have handsome advantages over textual representations:

- *locality*, information that needs to be used together or is logically connected is typically physically close together in diagrams (Larkin & Simon, 1987), reducing search time and the need for symbolic labels,
- *psychological emergent properties* due to well practiced perceptual inferences making working with diagrams easier than working with sentential systems (Koedinger & Anderson, 1994),
- *computational emergent properties* where geometric relations reduce the search space syntactically possible in the sentential systems, and,
- *structural constraints* where, for example, 'whole-part' relations in diagrams guide efficient knowledge organisation.

Searching for information from graphical representations is easier than with equivalent textual representations. This effect is a result of the mapping between the graphical system and what is represented. The explanation suggests a natural homomorphicity (Barwise & Etchemendy, 1995), or a tendency for the form of the representation to be similar to what is represented. There are however, problems in using diagrams according to Larkin and Simon (1987). A major weakness occurs when the limits of the representation are reached. When a representational system can support no more added graphical conventions, then the system may have to be supplemented with linguistic annotations, multiple level diagrams and other remedies. These added conventions make reading the diagram all the harder and detract from the processing efficiency of graphical systems.

A second approach refers to computational factors that describe advantages of graphical systems. One such research program identifies 'specificity' (Stenning & Oberlander, 1995) as the key factor.

Specificity accounts for a systems' benefit to the learner using principles that describe the representation's ability to communicate information. The complete set of information



that a system is able to express, determines an upper level of expression. This upper limit is the numerator in a ratio that is compared with the information communicated in a particular teaching and learning situation (the denominator). The approach simply states that the closer the two factors, the more specific and more effective the system will be. In general, graphical systems are thought to be more specific than sentential systems, and this is supposed to explain their pedagogical benefit. The specificity approach describes a way in which information is enforced in the most fitting representational system. The approach makes use of information or computational aspects of the domain and the representation.

The third approach describes ‘architectures’ of information processing. Architectures consist of processes and functional parts involved in diagram processing. Two research programs are characterized here, both emphasise the learner’s active attention to details in the diagram, and the value of the diagram to the learned outcome.

Paivio developed *dual-coding* theory (Paivio, 1971; Paivio, 1986; Clark & Paivio, 1991). This provides strong evidence to show the aptitude humans have in learning from visual stimuli. According to this work there are two independent coding mechanisms, one verbal and the other visual. Dual coding contends that pictures and words activate independent visual and verbal codes. The twin codes are additive so that information coded in both pictures and words is more likely to be remembered. Pictures are more likely to be coded both verbally and pictorially than words alone. Pictures are therefore more likely to lead to better learning than words because of the redundancy in the coding.

Comprehension from graphics involves imaginary manipulation of key constraining parts of the diagram (see Figure 2).

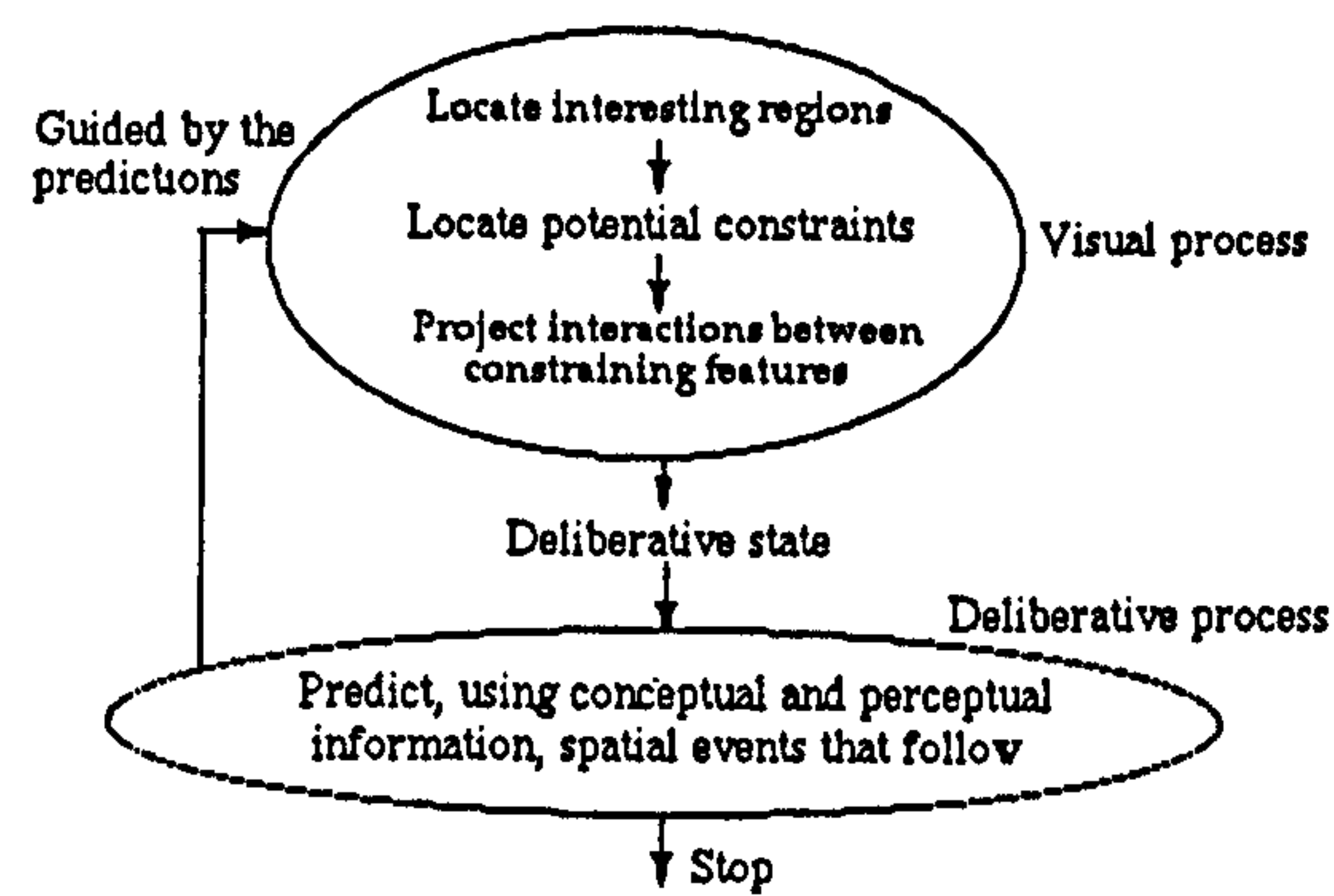


Figure 2: A model of visual reasoning  
(developed from Chandrasekeran & Narayanan, 1993),

Chandrasekeran & Narayanan (1993) believe that learner's abilities to predict the behaviour of a simulation is a result of exposure to previous cases of interaction with similar environments. A discrimination network is built out of a collection of frames, each representing cases and consisting of three parts: (a) visual conditions for the case which are verifiable through visual analysis of the diagram, (b) non-visual conditions which are verifiable through conceptual knowledge (semantics) associated with the representation, and (c) predictive events - the outcome of some interaction in the representation. Unlike Paivio's work (Paivio, 1971; Paivio, 1986; Clark & Paivio, 1991), this architecture is indifferent to underlying representation. There is no concern for the method of 'low-level' encoding. The learner accesses the diagram from memory and manipulates it there.

Both these architectural models have succeeded in replicating some of the features of the human processing of diagrams. The dual coding architecture particularly helps to explain why memory for information presented in diagrams may stay longer in long term memory than information presented in text. The most important implication for educational technology, lies in the precise way in which these architectures describe the process of learners' attention to diagrams. Both assert that the learner must make predictions about events from the interaction of constraining elements of a diagram. Both architectures assume the need for the learner to understand the intended meaning of the diagram lexicon.

These three approaches, the 'computational', 'architectural' and 'special purpose' explanations of diagram processing, are not mutually exclusive. For example, Sloman argues for a particular view of 'dual coding' described as 'redundancy'. He explains this point in a discussion over Glasgow's (1993) description of her own computational architecture of diagram processing. Glasgow points out that her use of rectangular data arrays do not necessarily encode information which could not be encoded propositionally. "It is not superior in *expressive power* (my emphasis), all it does is to provide rapid access to information via conceptually simple operations." Sloman argues that it is exactly this feature of redundancy in representation that makes for effective pedagogy. The memorization of the principles alone will not promote recall or effective performance and some consequences of simple axioms must also be memorized. Sloman is arguing for an architectural model that supports Paivio's representational scheme. Indeed, Glasgow's rectangular array representations are less expressive than a propositional alternative. This particular computational model suggests a role for 'specificity' in the architecture.

### *2.3 Translating between multiple representations.*

Languages allow demonstration and modeling of the domain, but also facilitate planning and execution (heuristic) stages in the problem solving (Goldin, 1987). Developing



proficient recognition and practice in these executive languages may require explicit training for the problem solver. Developing improved metacognitive skills in problem solving would improve performance from use of graphical representations. Indeed, executive languages used for processing visual notations may be different from those used for textual representations, and may require special attention. However these particular languages and their proficiencies are tangential to the role of representations used to illustrate the domain.

Multiple (external) representations are used to model and understand a domain of knowledge. While computer based courseware may use only one representational tool, the learner must still engage in translation to other languages to understand the topic. A learner with prior knowledge in a related but different field must translate the new material into terms that are understandable in the known language. Representations that are unfamiliar to a learner require redescription with known languages, before the new representations can become familiar and usable. Resnick (1982) used concrete versions of Dienes' blocks at the same time as describing their operation symbolically and found the strategy increased the learners' understanding of conservation. Wenger (1986) and Lewis (1991) both indicate that understanding complex systems hinges largely on an ability to use different models and map between them. Laurillard (1988) states that this 'translation strategy' makes explicit the relationships between one form of the representation and another, either semantic or symbolic or physical. Laurillard argues that translation is an example of a teaching strategy that gives the learner a rich experience of the conceptual environment.

Some topics of study furnish an opportunity to study multiple external representations that are all capable of illustrating the same topic area. These representations are the choices available to the courseware designer and the instructor.

#### *2.4 Discussion & implications for further study.*

There is great complexity in the use made of representations by the human learner. Representations that are created for demonstration by an instructor may end up acting as the primary language in which the model is accessed and manipulated in the learner's head. Whether a representation is internal or external, it may only be understood once the learner has translated the system into a language that was previously understood. Beyond these conscious languages of representation, the learner also has to store and retrieve information using languages of representation that may be unavailable to his conscious mind. These unseen representations may have direct consequences on the benefit that may be taken from representations that are available for inspection. There are several other languages of representation that help to modulate the processes involved. This complexity and speculation



over the exact mechanisms involved in learning from diagrams makes planning for empirical study quite difficult. However there are two encouraging themes to come from this review.

The first shows that there is substantial theoretical support for the development and testing of interactive graphical representations that effectively demonstrate the key elements of a topic area. Learning is an active process that demands the attention and effort of the learner in construction of some intellectual artifact (*e.g.*, Laurillard, 1995; Papert, 1993; Scardamalia & Bereiter, 1994). Learning with the use of graphical representations evidently requires the same active, attentive effort, perhaps more than for text. In cognitive models of problem solving (Glasgow & Papadias, 1995; Chandrasekeran & Narayanan, 1993), learners expend significant effort predicting the behaviour of the constraining features in a diagram. Interactive graphics that allow direct manipulation of the semantically important features, support this process for the learner. This interaction may promote the integration of new and continuously updated elements to the learner's mental model. Computer based implementations of diagrams can exploit explicit constraints of a representation that can only be implicit in a two dimensional line drawing. The activities that such systems offer to a learner include the exploration of those key constraints. The development of mental models is most likely to be of benefit when a systematic approach is taken to the exploration of new cases (Chandrasekeran & Narayanan, 1993). A continuously interactive diagram may support an infinite number of cases, and so a learner will still have the role of determining which cases to examine.

The second outcome responds directly to the first set of questions outlined in Chapter one. The 'specificity' research program provides a set of criteria (that are echoed by others), that makes strong claims about graphical systems. Those factors that are supposed to determine the benefit of graphical systems are also applied to choices between different graphical systems. This hypothesis provides the foundation for the studies that follow. The three models of diagram processing in model construction explain diagram benefit from different perspectives. These focus on the 'special purpose' nature of graphical systems, the 'computational characteristics' of diagrams, and the 'low level architecture' which encodes information in the graphical mode.

The decision to found the research agenda on one particular model considered the likelihood that results from investigating this model would provide the best information for implementation in a computer based learning system.

The architectural model gives an explanation of diagram use that, if proven, would advocate the use of diagrams, but would not help to understand features of diagrams or other representational systems that should be maximized for better learning. The architectural model provides some evidence that there is value in investigating diagram use, since it helps to explain why diagrams are useful to learners. The special purpose model of diagram use would show if proven, that diagrams for learning must be created by experts in a field. Such a

result would directly help courseware designers to focus their efforts in diagram creation and would ultimately help the learner. This kind of result, however, would not be amenable to automation. The special purpose model directly warns against looking for generic rules about diagram use for different topics. This leaves the computational model of diagram use.

The computational model of diagram use explicitly says that diagrams and other graphical representations are more vivid and tractable than other forms of representation. The model claims there are identifiable features of representations (expression and specificity), that help to determine why graphics are more successful as pedagogic aids, than other kinds of representations. The following analytic descriptions and empirical studies are driven by this claim. The computational model is the most likely basis on which to find precise description of graphical media that can be used either by an instructional media designer, or by the computer based learning environment, to make optimal decisions about the best media for the learner.

This review has presented explanations of diagram use without referring specifically to any particular topic. Investigating the claims of the model requires empirical studies conducted with learners developing a specific set of skills. The next chapter describes the topic used to test the computational model in the studies that follow.

**BLANK IN  
ORIGINAL**





### 3 Representations & Biases in Human Reasoning

#### 3.1 Syllogistic reasoning.

The previous chapter reviewed work in the cognitive and learning sciences that describes the processes of human learning with graphical representations. A framework characterized several intrinsic and variable features of graphical systems. Empirical and cognitive psychological approaches illustrated three ways to understand the benefit of the diagram for the learner. The account ends by selecting a computational account of diagrammatic reasoning as a working hypothesis for the project. Testing this hypothesis in an empirical context demands the application of the approach to a real learning situation in a particular topic area. For this work the topic chosen is syllogistic reasoning. The topic is small enough to easily illustrate, large enough to provide ample learning opportunities, lends itself to graphical representation and has several graphical methods available.

This chapter describes the domain of syllogistic reasoning, the representations used to help communicate it, and the problems learners have in understanding it. All the following experiments and analysis use syllogistic reasoning as the learning domain.

Many cognitive psychologists consider syllogistic reasoning to be the boundary of competence for humans. Learners' difficulties with syllogistic reasoning make the test of graphical systems a difficult one. Nevertheless, these difficulties also allow for substantial improvement from the right instructional strategy. The diagram systems illustrated in this chapter have all been used as instructional strategies for teaching syllogistic reasoning. The seven systems presented are: (a) Venn diagrams, (b) Euler circles, (c) Edinburgh Euler circles, (d) Johnson-Laird's mental model notation, (e) the predicate calculus, (f) Carroll's Game of Logic, and (g) TARSKI'S WORLD. Each system is distinct from the next in ways that are sometimes transparent and other times subtle and elusive. The descriptions of each system portray the smaller units or 'conventions' used by each system, what each lexical unit is intended to mean, and the transformations that are possible with them.

It is more than two thousand years since syllogistic reasoning was first written about (Aristotle, 336 BC, trans. 1983). For many reasons the complete skill-set turns out to be very difficult for most human problem solvers. A syllogism is a form of reasoning constructed from three simple statements that are modifiable to make variations in the form. Some syllogism forms have valid conclusions and others do not. The correct conclusion to a pair of premises is the valid conclusion or 'no valid conclusion' when the premises do not produce a conclusion. A statement (including 'no valid conclusion'), offered as a conclusion to two premises may be either valid or invalid.

Each problem consists of two premises each including a single quantifier and two terms. One term appears in both the first and second premise (called the middle term) and does not appear in a legal conclusion. The terms in the following syllogism are (a) qualitative representations, (b) diagrams and (c) graphical representations. Diagram is the middle term.

*Some qualitative representations are diagrams*  
*All diagrams are graphical*  
*Therefore, Some qualitative representations are graphical*

Premises and conclusions of syllogisms can be in one of four distinct moods:

- |                     |                         |           |
|---------------------|-------------------------|-----------|
| 1) All X are Y      | Affirmative universal   | (A);      |
| 2) Some X are Y     | Affirmative existential | (I);      |
| 3) No X are Y       | Negative universal      | (E), and, |
| 4) Some X are not Y | Negative existential    | (O).      |

Four possible figures result from arranging premises in different patterns. In each figure, A and C are the end terms, both occurring in the conclusion. B is the middle term, occurring in both of the premises.

- |       |       |       |       |
|-------|-------|-------|-------|
| 1.    | 2.    | 3.    | 4.    |
| A - B | B - A | A - B | B - A |
| B - C | C - B | C - B | B - C |

The first two figures are asymmetrical since the middle term is located in different columns in the two premises. The third and fourth figures are symmetrical since the middle term is located in the same column in both premises. Each premise can be in one of the four moods: A, E, I, and, O. Accordingly there are therefore  $4^3 = 64$  distinct forms. Different total numbers of forms are produced with different methods of counting. Recent views of syllogistic reasoning (e.g., Goldson & Reeves, 1992) indicate there are in fact 512 possible syllogisms. This number is arrived at since each pair of premises can be combined with eight possible conclusions, four different mood conclusions in the form A-C and four in C-A. The number of premise pairs yielding valid conclusions is independent of this issue. There are 27 syllogisms that yield valid conclusions.

Human learners find some of these forms very simple to answer. For other forms, learners often have no better than chance results of success (Johnson-Laird & Bara, 1984; Johnson-Laird & Byrne, 1991). Cognitive psychology has prioritized the goal to understand why different kinds of problems are more difficult to solve than others. In an early account, Johnson-Laird and Bara (1984) explain their motivation to understand the interactions between each candidate explanation.



A theory of syllogistic performance should at the very least account for the relative difficulty of different forms of syllogism for the figural response bias, and for the nature of the erroneous responses, including those of the type 'no valid conclusion' interrelating the end terms. The atmosphere effect can only account for some errors and not for those of the form 'no valid conclusion.' It is not intended to deal with relative difficulty or with the figural effect. The conversion theories certainly account for some errors and for some aspects of the relative difficulty of syllogisms, but they cannot explain either the figural bias or the erroneous 'no valid conclusion' responses. The Euler circle theories explain some aspects of relative difficulty but not the figural bias. (p. 2)

Understanding human performance in any topic area can help to influence and benefit instruction. Several research programs in learning and instruction emphasise the value of understanding the naive learner's model in a domain. Others similarly describe the range of misconceptions that tend to exist in a population (*e.g.*, Bowden & Walsh, 1994; Lybeck, Marton, Strömdahl, 1988). Models of human performance in syllogistic reasoning have developed from a need to understand human cognition and not to understand how to improve learning. Indeed some schools of thought appear to treat human reasoning as innate (Johnson-Laird & Bara, 1984; Johnson-Laird & Byrne, 1991).

### 3.2 Human performance.

Johnson-Laird's (1983) account of syllogistic reasoning uses "possible world" semantics as its foundation. This holds that the truth conditions of a sentence are determined by the set of all possible worlds that can be taken as models of that sentence. Using this principle he shows that valid reasoning is formally and empirically equivalent to the construction and manipulation of the possible worlds that can be created from interpretation of a sentence. The theory defines procedures for seeking counter examples and for revising models so far held in the head. The principal methods of explanation for the mistakes people make in solving syllogisms are the number and complexity of the models needed to solve the problem.

As we saw in Chapter 2, the general idea of mental models is that they share similar physical relational structure in the mind to the process being imitated (Gentner & Stevens, 1983; Johnson-Laird, 1988). In this case the mental models of reasoning represent different situations cast in the contextual worlds described in the syllogism premises. The theory assigns the reason for human failure to the increased load resulting from increased complexity in difficult problem types. Failure specifically results from having to consider and reject added possible worlds and to match each against possible conclusions. Errors are remarkably predictable in this topic area, even across cultures, which adds plausibility to the rigorous constraints proposed by Johnson-Laird's (1988) version of the mental model principle. That account required models to be strictly computable and for structures of the expert model to be identical to the structures of the states of affairs that the models represent.

The following three explanations and descriptions of error are sometimes called ‘error hypotheses’. Each is a candidate for explaining the predictable errors produced by humans: (a) atmosphere, (b) illogical conversion, and (c) the effects of belief. In each case a summary provides a list of errors that could be expected from an individual with the particular bias.

The intention of the analysis is to use this information as a form of diagnostic tool in a cognitive support tool for assisting the learning of syllogistic reasoning. The error descriptions provide a framework for describing behaviour to look for in the empirical work presented in Chapters 5, 6 and 7. Finally, these descriptions provide an understanding of mistakes that learners are likely to make in an instructional situation, and therefore pre-warn the instructor and provide an opportunity for diagnosing learners’ individual misconceptions.

3.2.1 The atmosphere hypothesis

Woodworth and Sells (1935) first described the atmosphere effect. Their analysis predicts the errors in syllogistic reasoning to be the result of a global impression produced by the premises, rather than because of strict logical deduction. Atmosphere is defined in terms of two dimensions; *quality* and *quantity*. Quality refers to the premise being either *affirmative* or *negative* and the quantity of the premise refers to whether the statement is *universal* or *particular*. The orthogonal pairings of values subsumed under quality and quantity yield the four moods described as: A, I, E and O. The atmosphere effect can be stated by two predictions. Whenever the quality of at least one premise is negative the quality of the most frequently accepted conclusion will be negative. When neither premise is negative the conclusion will be affirmative. The second principle referring to quantity states that whenever the quantity of at least one premise is particular, the quantity of the most frequently accepted response will be particular. When neither premise is particular the conclusion will be universal. The atmosphere hypothesis then states that the most likely error responses for each premise pairing can be shown as in Table 1 (below). The following two syllogisms have the same atmosphere, one has a valid conclusion and the other an invalid conclusion. The first is correct and according to the atmosphere hypothesis can be concluded because of the AA atmosphere of the premise pair.

All As are Bs  
All Bs are Cs  
All As are Cs

However in another case where the form is different although the atmosphere is the same, the conclusion is now invalid.



All As are Bs  
All Cs are Bs  
*All As are Cs*

The table below shows the moods one would expect as responses from premises presented to subjects. Entries in the column labeled “mood” are the moods of statements that would be expected from learners with the atmosphere bias, given the premise pairs in the second column.

The expected conclusions from learners with the atmosphere bias could also be stated as a set of production rules. If the premise pair includes an O, the expected mood would be O. If the premise pair does not include an O but does include an E, the expected mood would be E. If the premise pair does not include an O or an E, but does include an I, then the expected mood would be I. Otherwise the expected mood would be an A.

This theory cannot account for mistakes made where there is no valid conclusion or where there is a valid conclusion in another mood.

Table 1: Mood predictions from atmosphere hypothesis.

Mood	Premise pairs
A	AA
I	II, AI, IA
E	EE, AE,
O	OO, OA, AO, IE, IO, OI, EO

A restated version of the atmosphere hypothesis (Begg & Denny, 1969), corroborates the Woodworth and Sells (1935) account. The Begg and Denny (1969) account also shows that the results of illogical conversion (shown in the next section) can be explained by the atmosphere hypothesis. Begg & Denny (1969) take special care to emphasise that the mode of reasoning has not been proved by either the atmosphere hypothesis or illogical conversion position. Any definitive statement about what is really going on in the head must await research aimed at uncovering the processes involved in reasoning rather than in assessing the end product alone.

3.2.2 Illogical conversion

Chapman and Chapman (1959) describe an error hypothesis they call illogical conversion. Discussed also in Newstead (1990), the principle refers to the acceptance of the converse of either an A or O premise when not permissible and where the premise pair has a logically valid conclusion in another figure. Learners appear to misunderstand the difference

between premises with reversed terms (*e.g.*, All As are Bs, and All Bs are As), learners ‘convert’ the order of As and Bs to get a conclusion. This principle accounts for the main errors in: AA, AE, AI, IA, AO, and OA pairs.

Evidence supports illogical conversion which shows learners do better with syllogisms where the misinterpreted premises would yield the same conclusion as premises interpreted correctly (Wetherik & Gilhooly, 1990). Removing the possibility of conversion errors in these studies increases performance, indicating that the error contributes to poor performance. Learners also do better with premises worded to explicitly avoid conversion (Ceraso & Provitera, 1971). Premises that are empirically true as All Bs are As, are less likely to be converted to an All As are Bs premise. For example, the premise, All Sharks are Fish is less likely to be converted than the premise, All Fish are Sharks.

The following syllogism has a valid conclusion:

*All As are Bs*  
*No Bs are Cs*  
*No As are Cs*

When the first two terms are switched around, as shown in the next example, so that All As are Bs becomes All Bs are As, subjects will most often give the conclusion, No As are Bs (as above). The new syllogism with correct conclusion may however have only a restricted negative conclusion as shown.

*All Bs are As*  
*No Bs are Cs*  
*Some As are not Bs*

The error lies in the misunderstanding of the nature of the A mood. The illogical conversion hypothesis only relates to syllogisms where there is another figure including the same combination of premise moods where the accepted conclusion would have been correct.

3.2.3 *The effects of belief*

The most pervasive group of problems in reasoning are related to the effects of what learners believe to be true about the world. The associations people have with terms and the relationships between terms can affect the conclusions they are willing to accept as possible and the processes used to reach those conclusions. There are according to Oakhill, Garnham, and Johnson-Laird (1990), three principal ways in which beliefs can affect reasoning: (a) they may distort the interpretation of the premises, (b) they may influence the deductive process, biasing which conclusions are reached, and (c) they may be used to filter out unacceptable conclusions that are produced by the deductive process. Some argue that the reasoning process is independent of belief and can therefore only affect the interpretation of premises.



Wason and Johnson-Laird (1972) account for learners' misinterpretations of premises based on beliefs. They test learners with statements of the form, "Some As are Bs" where the correct interpretation would be that at least one but possibly all As are Bs. In their experiment half of the learners received semantically 'universal' some statements suggesting universal interpretation, for example, some beasts are animals (empirically by most definitions at least, all beasts *are* animals). Others received 'non-universal' some statements such as, "Some books are novels," for which only some books are empirically novels. To check interpretation of the premises, learners were given diagrams from a graphical system used to demonstrate syllogisms, called Euler circles. Two variations of this system are presented later in this chapter. Subjects were asked to identify the diagram that best represented the statement, "Some As are Bs."

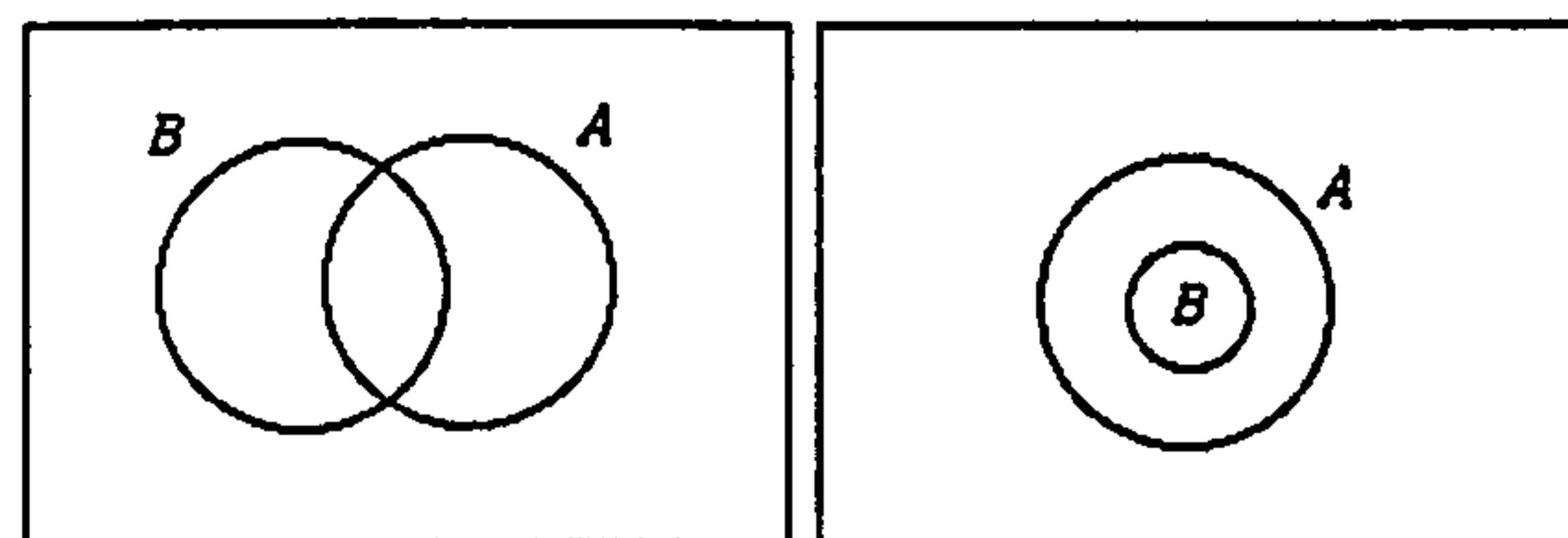


Figure 3: Wason's diagrams.

The study assumes the first diagram (with two intersecting circles) properly represents the statement, "Some As are Bs." Results of the study showed that 75% of respondents that were exposed to the misleading premises, chose the second diagram. This was compared with only 25% of learners making the same mistake (choosing the second diagram) with the non-universal statements. Wason and Johnson-Laird's (1972) experiments are intended to show that while the form of the syllogism is held constant and the linguistic content of the premises are altered in potentially misleading ways, subjects are significantly influenced by their beliefs about the terms. It is not clear these experiments are conclusive for several reasons. All these reasons refer to the representations and their interpretation by experimenters and subjects.

The first problem involves the learner's understanding of the graphical systems used. Experiments reported in this project (Chapters 6 & 7) illustrate plainly that before any training, learners have quite well defined preferences and assumptions about the meanings of alternative diagrams. These assumptions do not correlate well with standard semantics and allowable transformations from the original systems. Learner's understanding of the diagram systems probably did not influence their choice of diagrams in this study.

The second problem again involves the diagrams presented to the learners. The Euler circles used, represent only two of the four possible worlds that properly represent situations for which the statement, "Some A are B," would be true. The four types can be demonstrated by four example situations with the form, "Some As are Bs": (a) Some bachelors are

unmarried men, (b) Some books are novels, (c) Some artists are bee-keepers, (d) Some cod are fish. According to the traditional Euler circle system, each of these statements could be represented with one of the following diagrams in Figure 4. All diagrams are therefore *legitimate* representations for the general form, Some X are Y. The circles in Figure 4 are labeled with the first letters of the terms in the examples given above. The first diagram for example represents the statement, “Some bachelors are unmarried men.” Because the factual (belief related) information in this premise is actually tautologous, the two term circles collapse to form just one circle. Each of the other diagrams show the other three kinds of potentially belief altering situations in which the premise would also be true.

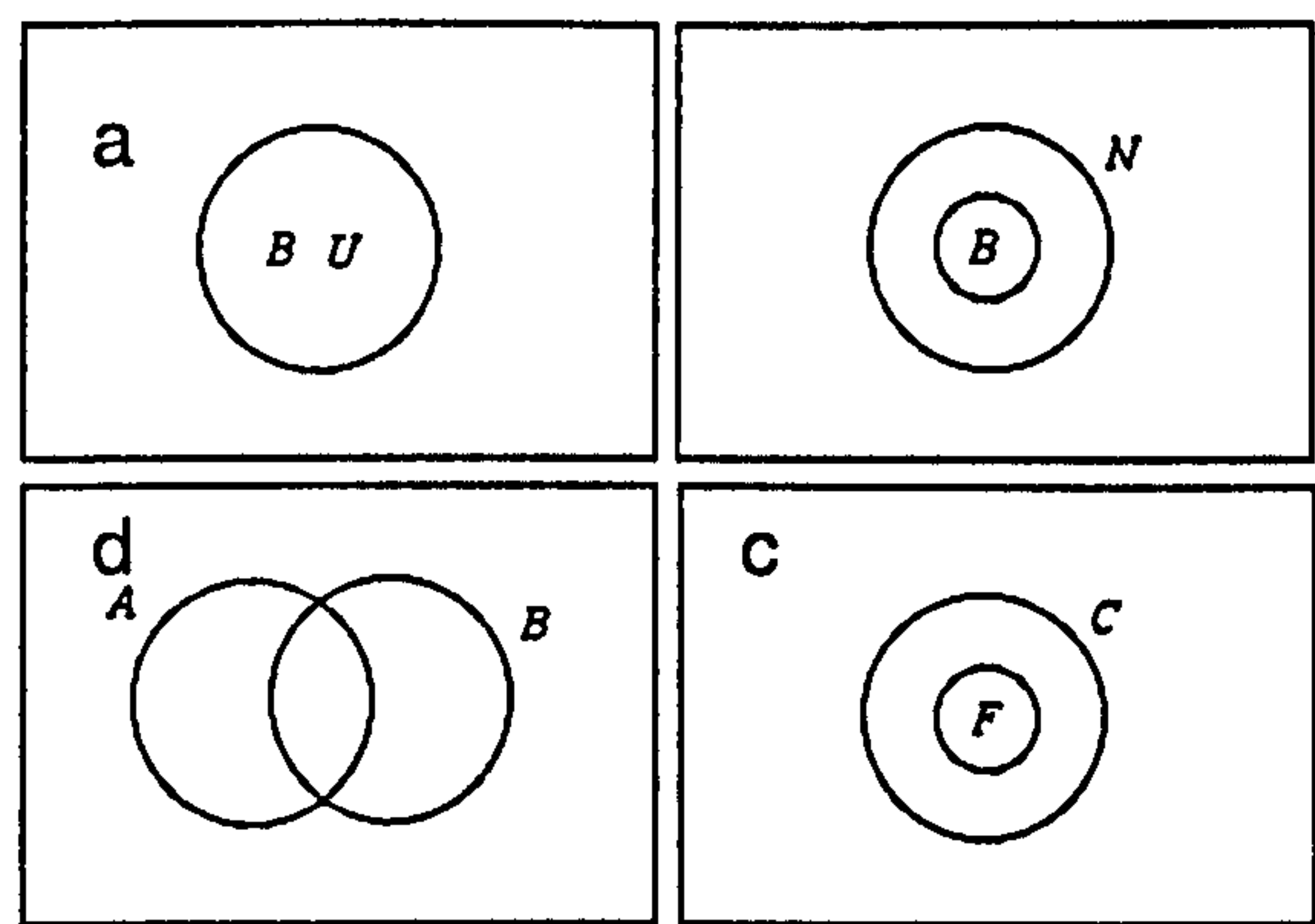


Figure 4: Conditions for Some X are Y.
   
 Note: The labels refer to the four types of premise in the text, where: *B* = Bachelor, *U* = Unmarried man, *B* = Book, *N* = Novel, *A* = Artist, *B* = Beekeeper, *C* = Cod, and *F* = Fish.

Other circle based systems have equally well defined semantics and could have been used in place of Euler. Some diagrams in these systems have the same appearance as the diagrams from the Euler system but have different meanings. While diagrams may help natural deduction when they are properly understood, the alternative semantics for these alternative systems forces the conclusion that no one diagram system can be regarded as natural.



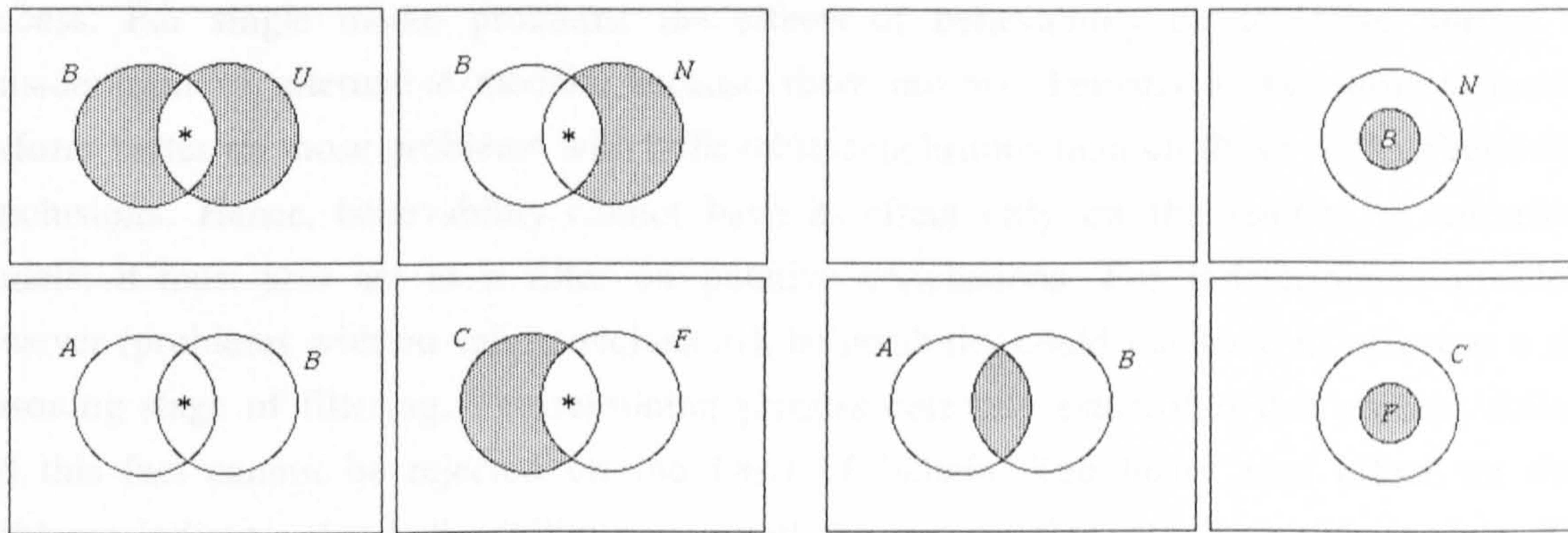


Figure 5: Venn & Euler illustrations of the four contingent states which may be true of Some X are Y.

*Note:* The four diagrams on the left illustrate Venn diagrams and the four on the right are Euler circles.

The third and final problem with this study concerns the experimenters' use of the Euler system. Strictly speaking, neither of diagrams offered to learners in the experiment are valid representations (with the Euler system) of the statement, "Some X are Y." The circle system illustrates possible situations. No single situation is sufficient (in Euler) to conclude the, "Some X are Y" statement. Results of the Wason and Johnson-Laird (1972) study may just as likely reflect the correct identification of a *situation* for which the premise type is true. Learners choosing a particular diagram may not have mistaken the intended meaning of the premise, but instead, made an accurate choice of diagram to show the statement provided in the material.

The study does show that learners choose diagrams differently given premises that have different empirical situations. Beliefs, however, can also affect the process of deduction. Revlin and Leirer (1978), for example, provide results that show: "Even when the logically valid conclusion is not altered by conversion, if logic and belief point to a different conclusion, subjects were less accurate than when logic and belief concurred" (p.51-81).

The deduction process can be closely monitored in experimental designs that allow subjects to create their conclusions instead of choosing from multiple options with distracters (Oakhill, Garnham, & Johnson-Laird, 1990). In their study (1990), subjects performed well in determinate problem cases where suggested conclusions were empirically unbelievable, but poorly when the conclusions were definitionally unbelievable. Empirically false statements are just contrary to general knowledge (*e.g.*, Some of the athletes are not healthy), definitionally false statements are false because they violate a definition (*e.g.*, Some millionaires are not rich). Their work (1990) suggests an influence on the deductive process itself since neither of the forms in the investigation could have different conclusions as a result of misinterpretation of premises. Oakhill and colleagues further develop the argument that the effects of belief are not restricted to either the deduction process or the filtration



process. For single model problems the effects of believability cannot arise during the consideration of alternative models, because there are no alternatives. Yet subjects tend to perform better on those problems with believable conclusions than on those with unbelievable conclusions. Hence, believability cannot have its effect only on the search for alternative models; it must also act as a filter on putative conclusions. For indeterminate problems however (problems with no valid conclusions), believability could not have its effect at a post reasoning stage of filtering. The reasoning process correctly establishes that nothing follows and this fact cannot be rejected on the basis of beliefs. The belief bias effect on these problems indicates that believability can curtail the examination of alternative models. This finding cannot be reconciled with the idea that the effects arise only at the stage of filtering conclusions.

### 3.3 Representations of syllogistic reasoning.

Some diagrams have been used to illustrate aspects of the syllogism in the previous sections. This part describes systems for representing syllogistic reasoning with careful attention to their semantics and syntax. The diagram systems outlined here are described in abstract form. The level of detail is enough to explain how each works but does not describe the underlying logic in a formal way. Other research provides formal descriptions for the systems (*e.g.*, Shin, 1991; Allwein & Barwise, 1996). Each system described here includes the meaning of symbols used and the processes available to manipulate those symbols. The symbols are called “conventions” of representation. Each system is a perfectly legitimate alternative, but the symbols must be understood as their inventor intended. The process of manipulating the symbols to reach a conclusion is called the “algorithm” for solving the problem. This is a way to clearly describe the process in small steps.

The systems were developed at different periods of history, sometimes separated by hundreds of years. Each inventor claimed their new system was an improvement over what it was intended to replace. For example Carroll (1958) said:

My Method of Diagrams resembles Mr. Venn's, in having separate Compartments assigned to the various Classes, and in marking these Compartments as occupied or as empty; but it differs from his Method, in assigning a closed area to the Universe of Discourse, so that the Class which, under Mr. Venn's liberal sway, has been ranging at will through Infinite Space, is suddenly dismayed to find itself cabin'd, cribb'd, confined, and in a limited Cell like any other Class! (p.12)

Nonetheless, each of the representations for the syllogism share certain characteristics: (a) they are predominantly graphical which is to say they rely on specialized spatially oriented notations, (b) they are external in that before internalization they are the subject of *conscious* deliberation and (c) they are all monosemic -- designed to be interpreted in only



the way intended by the inventor. These representations meet most of the criteria for *homomorphic* systems laid out by Barwise (1994). This means their conventions share the same form as their referents in a very strict way (Chapter 2). The similarities contrast with differences between systems discussed in the final section of the chapter.

The syllogism can be demonstrated with many different notations and it is possible to create new ones too. The account here is not a comprehensive account of all possibilities. Seven systems are shown here: (a) Venn diagrams, (b) Euler circles, (c) Edinburgh-Euler circles, (d) Johnson-Laird's mental model notation, (e) the first order predicate calculus, (f) TARSKI'S WORLD, and finally, (g) Lewis Carroll's 'Game of Logic'. Four of these systems have been developed as software. Each application uses the graphical system illustrated here. The functionality of each software system is quite different; sometimes different from the abstract systems shown here. The functionality and operation of these software applications is described in Chapter 5.

Not all seven of these systems continue to take a role in the following chapters. The descriptions of Venn, Euler, Carroll and TARSKI'S WORLD are important for the following studies and analysis. Descriptions of the other systems are not important to the thesis argument and may be skipped. A standard problem is solved in each description (All B are A, Some B are C  $\rightarrow$  Some A are C). The description is given in two parts. The first provides a portrait of the lexical items and their meaning. The second describes the process of manipulating the diagrams to reach a valid conclusion.

### 3.3.1. Venn diagrams

In the Venn system, three intersecting circles occupy a rectangle. The rectangle counts as the universe of discourse, the superset of all possible terms. Each of the circles represents one of the terms in the syllogism. Shading shows there is no possibility for an individual to occupy the region. A star or linked chain of crosses show a region where individuals must exist. If a premise states that an individual exists within an ambiguous area, where there are two possible regions, then the cross is placed on the line between the two. The system can express partially specified individuals, for example things which are "A and B and either C or not C" (Figure 6b demonstrates this partially specified individual). Systems that support representations of ambiguous situations cannot be called *homomorphic* according to the criteria set out by Barwise and Etchemendy (1995).

Figure 6a shows the syllogism, "All B are A, Some B are C  $\rightarrow$  Some A are C." The shaded lunule represents the non-possibility of, Bs that are not As, since All Bs are As. The star represents the known existence of a C that is also a B.



Conclusions are formed from reading the relationship between the A and C terms from the diagram. Since the star is in the intersection of the A and C circles, an existential conclusion is possible, Some As are Cs.

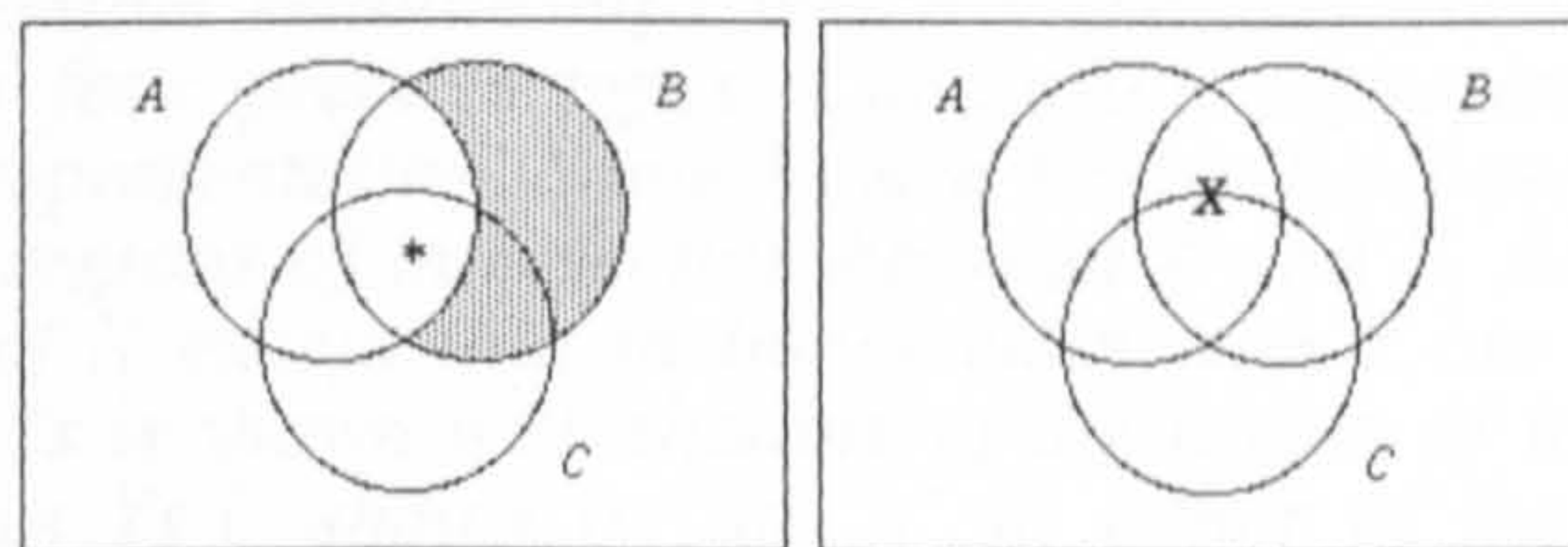


Figure 6a and 6b: Venn showing All B are A, Some B are C. (6b shows a part specified individual).

As the other systems used in the project, the Venn system can be used with variations of the notations. For expressing disjunction, sometimes a chain of 'x's are placed across boundaries, shading can be replaced by ticks, and stars may take the place of 'x's. For the purposes of this project we assume this standard representation of either stars or crosses and shadings. The mechanism for each variation remains the same. The semantics of the lexicon used in the rest of this project are shown in the next section.

#### *Conventions of representation*

Representational conventions used in Venn are restricted to four primitive objects, (a) the closed curve which is usually circular, (b) the rectangle, (c) shading and (d) the cross (or chain of crosses). The closed curve relies on the principle of containment to show the boundaries of a named set. The rectangle uses containment to show the boundary of the universal set. The rectangularity distinguishes the universal set from other closed regions. Shading represents the known emptiness of a set. The cross represents the known non-emptiness of a set. An extended description of Venn's semantics based on a situation theoretic account can be found in Shin (1991). Shin used the chain of 'x' notation and extends the basic Venn system to cover similar quantified problems with more than three terms.

#### *Algorithm for problem solution*

Solving a syllogism can be thought of as the unification of two well-formed Venn diagrams. Specifically the sub-set is determined by the rules of the syllogism. Each of the two diagrams must have two terms, and one term will be shared between two diagrams. The algorithm for this restricted version of Venn diagram unification is as follows:

- 1) *Set up initial premise diagrams one for each premise. Since there are only four premise types and only one diagram for each type there are four options at this stage,*



- 2) Register the B term circle from the two premise diagrams ensuring that there is intersection between the end terms;
- 3) Resolve any existential terms which are disambiguated through the registration process. That is any 'x's which were on arcs between regions and where one of those regions has been precluded by the unification process should be resolved to the unshaded region, and
- 4) Read the conclusion by looking for the relationship between the A and the C terms. There are four possible relationships which there may occur between A and C which correspond to the four premise types. Conclusion representation however is not the same as premise representation. Some Xs are Ys can be shown when the 'x' is in either of the ambiguous regions of intersection between A and C. All Xs are Ys can be shown when all regions of X except that in intersection with Y are shaded and are therefore empty. No Xs are Ys is shown with shading of the region of intersection between X and Y. Some Xs are not Ys is shown by an 'x' in either of the ambiguous regions of X which are intersected with Y.

There are several other accounts of the Venn algorithm. The first is in Venn's original text (1894). Sun-Joo Shin's account is very detailed and was the first to describe Venn's system in formal semantics and syntax (1991). Atsushi Shimojima's account (1996) focuses on a formal account of the constraints in Venn's system. Others refer to Venn's system in passing (e.g., Barwise & Etchemendy, 1991; Gardner, 1958).

### 3.3.2 Euler's circles

There are at least two variants on the Euler circle system. Much of the psychology of reasoning literature uses the more traditional Euler circle system (e.g., Ceraso & Provitera, 1971). A more novel interpretation of Euler (Stenning & Tobin, 1993; Stenning & Oberlander, 1995) is demonstrated in the following section.

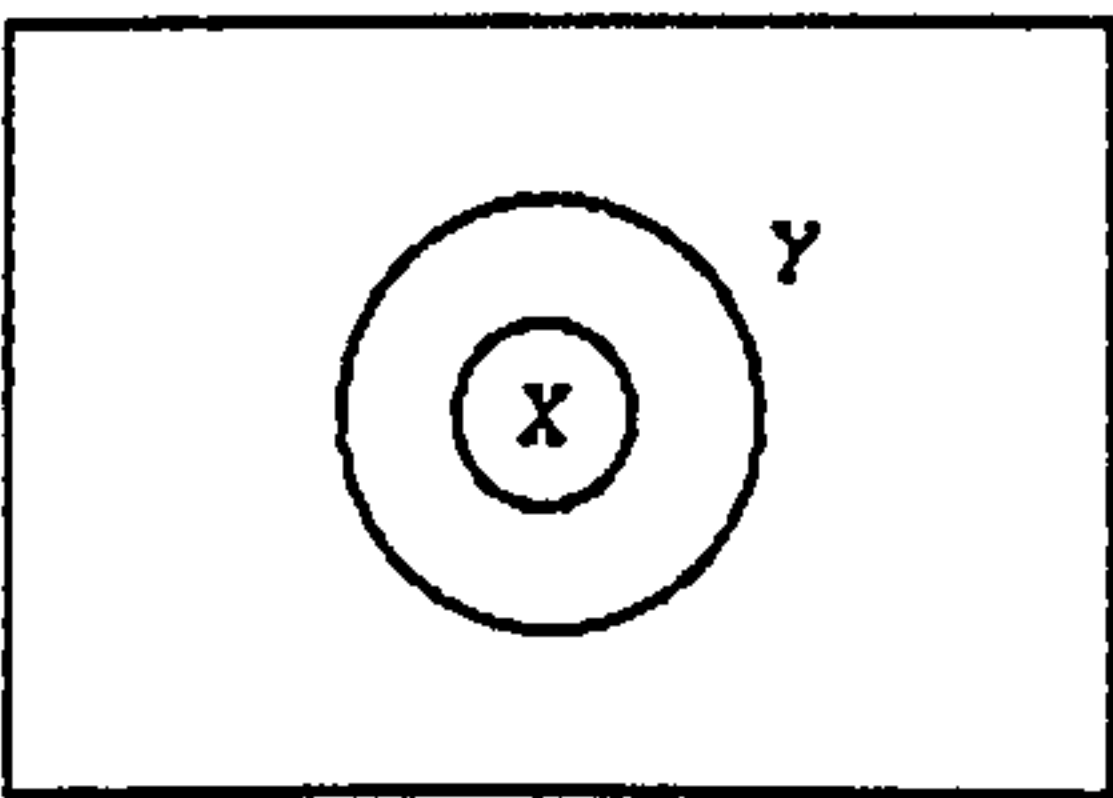
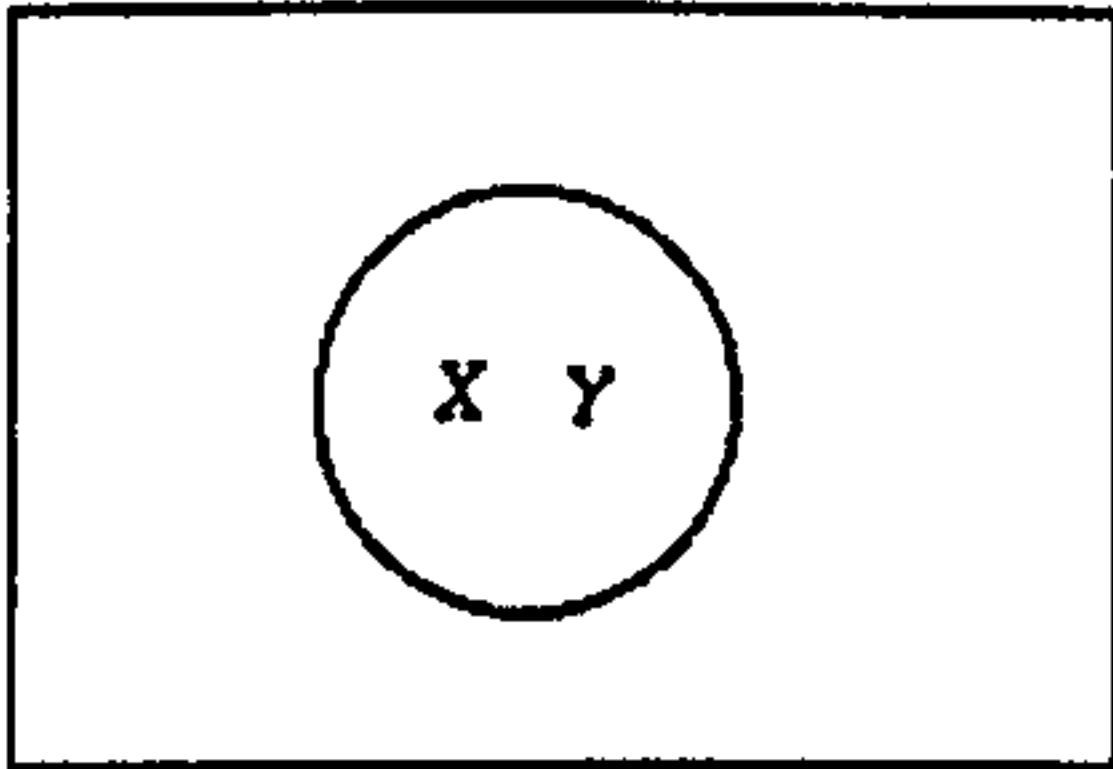
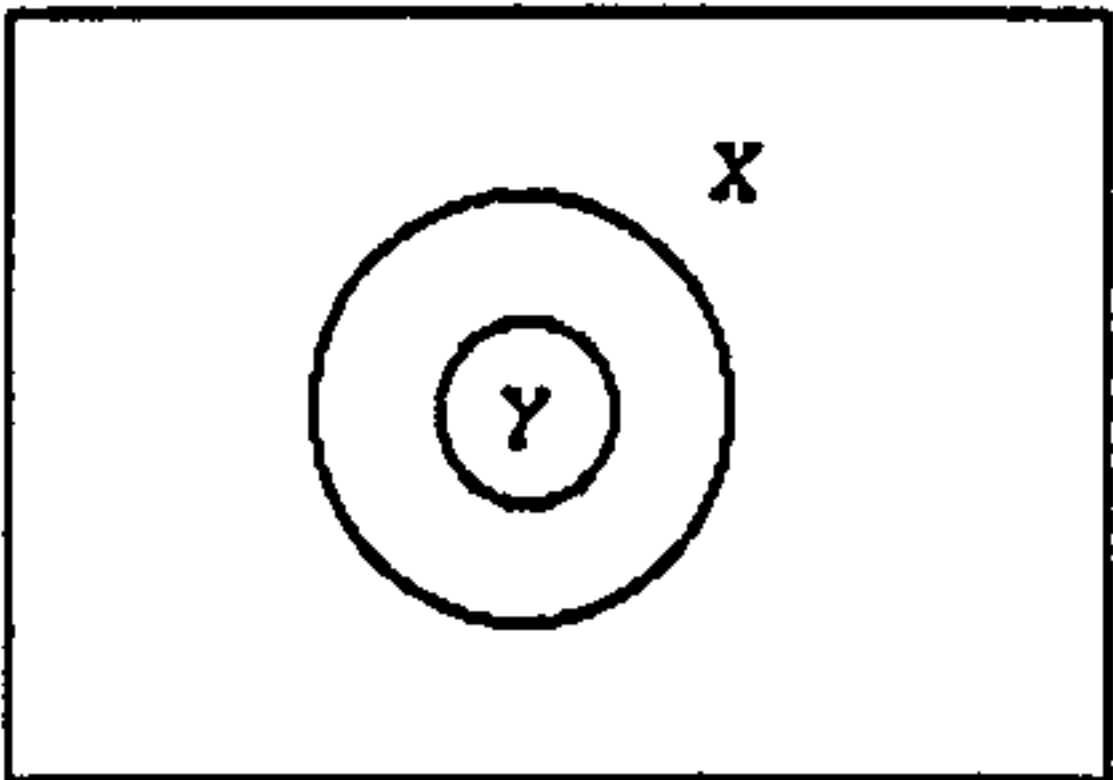
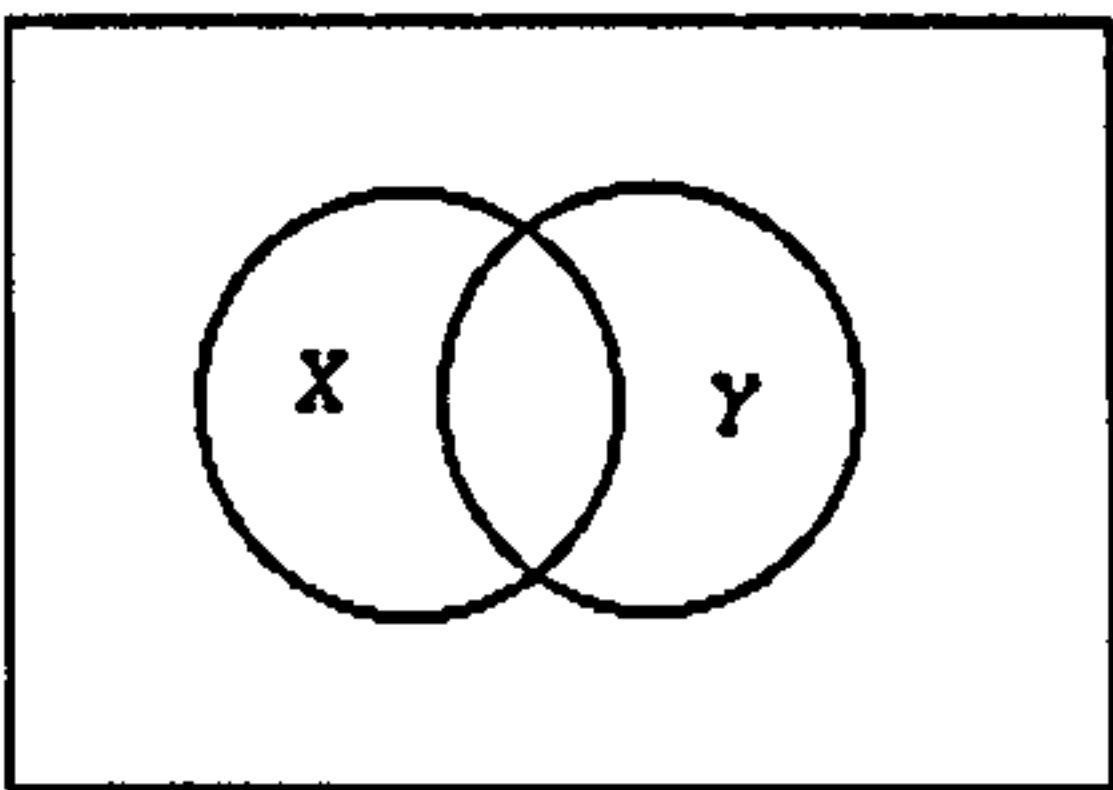
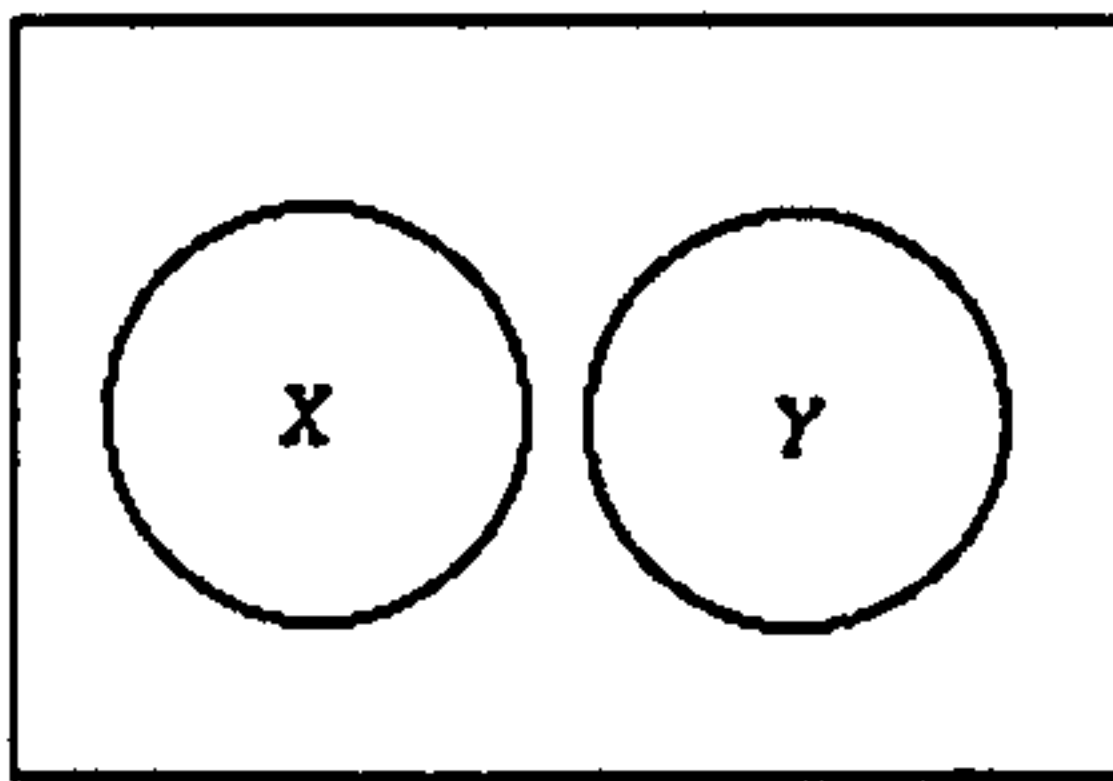
The old system begins with situations that exist between two terms; showing those situations with two circles. The circles may be either: separated, intersecting, the first contained within the second, the second contained within the first, or both circles identified with each-other. These relations are known as the Gergonne relations and are shown in Table 2 below. One diagram (or situation) is a valid representation for more than one statement. For example, two intersecting circles illustrate a world in which the statement, "Some Xs are Ys" and the statement, "Some Xs are not Ys" are both true.

This kind of system means that there are many representations for a single premise. This leads to a proliferation of diagrams for solving any single problem. It has been argued (Ceraso & Provitera, 1971) that this older Euler circle system can give an indication of the complexity of individual problems. Complexity was calculated by counting the number of possible A-C and C-A combinations for the syllogism. It is possible according to Ceraso and Provitera (1971) that when presented with written or spoken syllogisms, people *see* only one diagrammatic representation in the head and only one model. These diagrams may be the first model of many that are required to prove the syllogism. If the learner is unable to read a conclusion from this first diagram, he may quit and answer that there is no valid conclusion.



According to this theory, subjects fail when there is more than one Euler representation of the premises, and where a conclusion derivable from the first diagram was not derivable from further diagrams that were not observed.

Table 2: The Gergonne relations.

	<i>All X are Y</i> <i>Some X are Y</i>
	<i>Some X are Y</i> <i>All X are Y</i>
	<i>Some X are Y</i> <i>Some X are not Y</i>
	<i>Some X are Y</i> <i>Some X are not Y</i>
	<i>Some X are not Y</i> <i>No X are Y</i>

Solving the syllogism with the traditional Euler circle system relies on the combination of all possible diagram types for each premise type. Any valid conclusion must be true in all the possible conclusion diagrams. In the left column of Figure 7 are the representative diagrams for each premise. The left column diagrams (labeled  $A_{1\&2}$ ) are the two legitimate possibilities for the premise, All Bs are As. The second column (labeled  $B_{1-4}$ ) shows the four diagrams for the premise, Some Bs are Cs. The third (labeled  $A_1*B_{1-4}$ ) and fourth columns (labeled  $A_2*B_{1-4}$ ) show the conclusion diagrams. The third column is the result of combining the first diagram of first premise with the four diagrams of the second premise. The fourth column is the result of combining the second diagram of the first premise with the four diagrams of the second premise.

There are therefore eight conclusion possibilities. In two cases there is more than one possible diagram for one premise diagram combination. Multiple diagrams are possible in some cases because there is insufficient restriction in the algorithm to determine just one diagram from its parents. In this syllogism there are twelve possible diagram combinations for the conclusion. In all the conclusion diagrams, the relation “Some As are Cs” is preserved. This check on the conclusion diagrams establishes the statement as a legitimate conclusion.

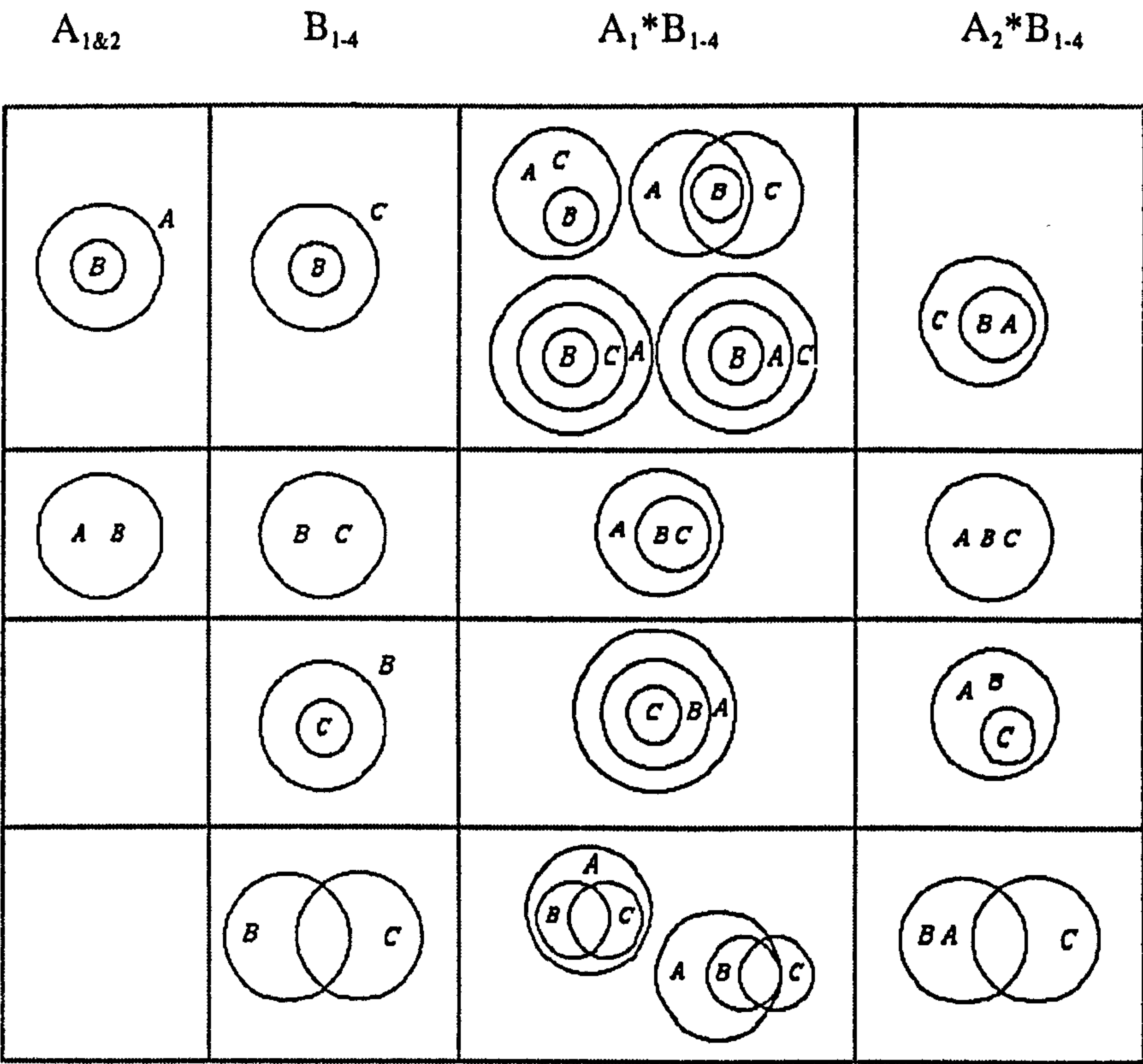


Figure 7: The older Euler method.

The older Euler method. Possible A-C relations for the syllogism All B are A, Some B are C. The relation Some As are Cs is always preserved in conclusion diagrams.

Conventions of Representation

Within the older Euler system there are very few conventions. The closed curve represents the set. Containment shows membership, intersection shows commonality, separation shows non-commonality. There is no shading and no existential annotation so that all regions are assumed to be non-empty. The universal set is shown by a rectangle that contains all other closed curves.

### *Algorithm for solution*

The algorithm for the older Euler system is simpler than for Venn. In practice, the algorithm produces more steps than for Venn and is implausible as a way to support the reasoning process in the head just because of the many diagrams that it generates. The difference between Venn and Euler indicates a trade-off between the complexity of the diagrams and the number of steps in the algorithm. The algorithm in this case has simple steps that repeat many times for multiple model problems.

- 1) *The premise diagrams are constructed and there will be as many as four diagrams for one premise and so between two and eight diagrams for a pair of premises;*
- 2) *A pairwise registration is made for each of the possible combinations. This may result in anything between one to sixteen possible combinations, although certain diagram combinations generate more than one solution (there may be more than 16 diagrams), and*
- 3) *The conclusion is derived by positing one of the possible relations between A and C and checking it against all of the resultant combined diagrams to see if that conclusion is true in all cases.*

While the algorithm for Euler is simpler than Venn's, particularly in the conclusion derivation phase, the number of steps can be significant when there are many premise diagrams. The older Euler system involves many repetitious (but simple) steps.

### *3.3.3 Edinburgh Euler*

The Edinburgh Euler system is an improvement on the traditional Euler system. The new system is simpler because it reduces the number of diagrams that result from combining the premises. The new Euler system avoids this explosion of diagrams by adding new conventions and new constraints. Fewer steps are required to solve the syllogism. These steps involve several different checks and decision points that contribute to a more complicated algorithm.

In the older systems the four possible premise types: A, I, E and O could be represented by 5 possible circle configurations: A by 2 configurations, I by 4, E by 3 and O by 1. In the newer system with its added conventions there is a one-to-one mapping from the premise configurations to diagrams. There are four diagrams and four premises. One diagram represents one premise type. Figure 8 shows the newer Euler representation of the syllogism, "All Bs are As, Some Bs are Cs  $\rightarrow$  Some Cs are As."



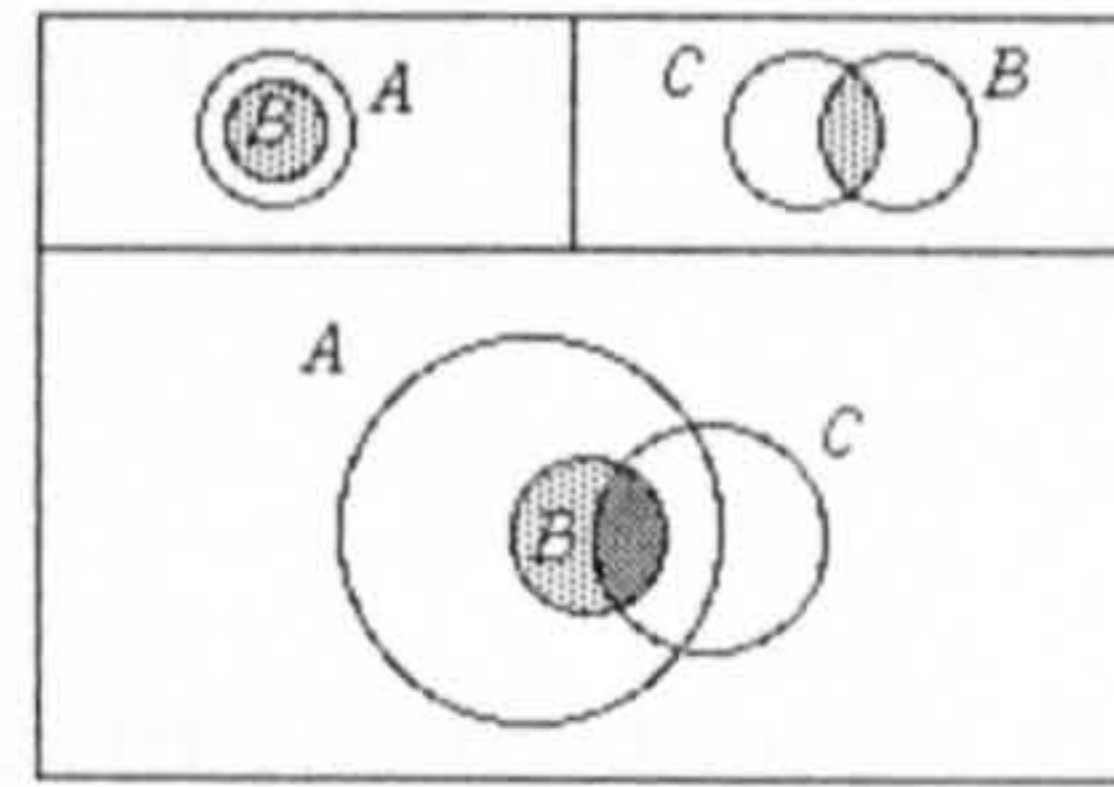


Figure 8: The Edinburgh Euler method.

The Edinburgh-Euler system represents each premise by a single diagram and combines the two premise diagrams in a clear and systematic method. The problem represented in Figure 8 would have required four possible diagrams for the first premise.

The shading convention reduces the number of diagrams required by stipulating the existence of individuals in the premise diagram. This reduces the situations that can be represented. The older Euler circle system allowed different diagrams for statements: (1) Some As are Some Bs, (2) All As are All Bs, (3) Some As are All Bs, and (4) Some and only Some As are Some and only some Bs. In the Edinburgh-Euler system only the last of these (Some and only Some As are Some and only some Bs) is allowed.

The next chapter develops a system for calculating the match between the information that a representation can illustrate and the information in a domain being studied. The new Euler system is a better match than the older system since it is able to express less information, but enough for the domain to be taught. The older system sometimes demanded as many as 16 separate diagrams for a conclusion. There is only ever one conclusion diagram with new Euler.

Although Edinburgh-Euler is the combinatorially simpler system, the process of registration may not be as obvious to the learner. The process is developed fully by Stenning, Cox and Oberlander (1992). It begins with development of characteristic diagrams of each premise. In registration the goal is to combine the middle term circle with as many regions as possible relating the end terms. Any shaded areas that are left in the registered diagram are compared with the parent premises. If any shaded area remains non-intersected from the premises, and no more than one premise was negative, then there is a valid conclusion. The formulation of the conclusion depends on the quantification of this non-intersected area. This algorithm is expanded below.

### *Representational Conventions*

The conventions used in Edinburgh-Euler include: (a) the closed curve, (b) shading, (c) intersection and (c) the rectangle. This system has more conventions than the older Euler or Venn. Containment shows membership of a set although there is no non-empty set assumption in this system. Shading is used in the opposite way to Venn, to show known existence within a region (shading in Venn shows known non-existence within a region). Circle separation is used to show non-commonality of sets. The membership of any set has to



be explicitly stated by the use of shading. Circle intersection shows possible, but not necessary existence of any individual. The rectangle shows the universal set.

#### *Algorithm for solution*

The algorithm consists of eight identifiable stages:

- 1) *form characteristic diagram for each premise (there are only four options for each premise);*
- 2) *register the B circles of the characteristic diagrams of the premises and arrange A and C circles with most types, that is regions and intersections consistent with the premises;*
- 3) *if no shaded region from a component premise remains non-intersected then exit with the no valid conclusion response however if there is one, then it is the critical region;*
- 4) *if such a region does exist but both premises are negative, then exit with No Valid Conclusion response;*
- 5) *formulate conclusion by taking the description of the individual represented by the critical region of the diagram for example,  $A \neg BC$ ;*
- 6) *eliminate the B term from this description for example, AC;*
- 7) *existentially quantify the remaining description for an existential response, Some A are C, and*
- 8) *if the critical region is circular and labeled by an end term, then it is the subject of a universal conclusion, All As are Cs, if not there is no universal conclusion.*

The reduction of representational complexity in the diagrams may have been at the expense of increased complexity in the algorithm. This implies a trade-off between the dynamic and static aspects of the representation. This in turn may have an implication on the choice of system for a learner.

Instruction in any discipline may choose to focus on performance or on understanding. The goals of instruction will influence the choice of instructional strategy. Diagram systems with simple conventions and several repetitious steps may favor accurate performance but provide less opportunity for understanding. Diagram systems with complex algorithms and simple diagrams may be better for illustrating the domain but difficult to use to solve problems.

#### *3.3.4 Johnson-Laird's mental model notation*

Johnson-Laird's system of notation for the syllogism (Johnson-Laird, 1983; Johnson-Laird & Bara, 1984; Johnson-Laird & Byrne, 1991) is as much a claim about the plausible internal processes involved in syllogistic reasoning, as it is a representational system for teaching the skill. There are three stages in syllogistic problem solving for Johnson-Laird: (a) a mental model of the first premise is constructed, (b) the information in the second premise is added to the mental model of the first premise, taking into account the different ways in which this can be done, and (c) a conclusion is formulated that expresses the relation between the A and C terms. This relation must hold in all models of the premises.

If in stage two the generation of all possible models is not completely satisfied then stage three will likely generate conclusions the truth of which is not guaranteed. The actual implementation of the mental model construction requires the imagination of an arbitrary set of symbolic tokens for each of the premises.

The notation is described by Johnson-Laird & Byrne (1991, p119) and uses a new line to represent different kinds of individuals that are inferable from a premise. Premises may allow more than one new kind of individual. In fact the number is arbitrary, although reasoners are likely to represent only as many as necessary to illustrate the possible kinds. To illustrate the premise, "All bakers are athletes," the reasoner should use a model represented by the following notation.

```
a [b]
a [b]
...
```

The bakers are represented by the "b" tokens and the athletes by the "a" tokens. The square brackets around the "bs" indicate that the set of bakers have been exhaustively represented in the model, *i.e.*, members of the set cannot occur anywhere else in the model except as individuals that are not bakers. Individuals known not to be bakers are shown with the notation, "¬b".

The next step is to add the second premise, "some bakers are carpenters" to the model. Since the "b"s are already exhaustively shown and no more may be added, the "c" must refer to one of the "b"s already indicated. The "c" is added to a line with a "b" that already exists. This leaves the following model that shows there is at least one "a" that is also "c" and therefore allows the conclusion, Some As are Cs, or some athletes are carpenters.

```
a [b]
a [b]c
...
```

The three dots, sometimes called "anaphora," show that other types of individuals are allowed for, but have not yet been made specific. These other individuals should not affect the truth of the first premise. In the model immediately above, there could be other individuals that are athletes, but are neither bakers or carpenters. The possibility of such an individual is not prohibited by the model above and therefore prevents the universal conclusion that all athletes are carpenters.



The first example was a particularly simple syllogism that 76% of subjects in Johnson-Laird and Byrne’s studies (1991) were able to answer correctly. To demonstrate the mechanism more fully, another three-model syllogism may help.

All builders are architects  
No builders are choirists.

This syllogism is more complicated, but yields the following model. The builders are exhaustively represented with respect to the architects, and the “a”s and “b”s are exhaustively represented with respect to the “c”s. The model prohibits any other “b”s that are not “a”s, as well as any “b”s that are also “c”s.

[a [b]]  
[a [b]]  
          [c]  
          [c]  
...

This model represents a scenario that could support the conclusion,

No choirists are architects.

This is the most common error with this problem but the following model refutes the conclusion. The new “a” can be added to one of the “c” individuals since, the earlier model introduces no such restriction.

[a [b]]  
[a [b]]  
a      [c]  
          [c]  
...

This new model supports the conclusion:

Some choirists are not architects.

However this conclusion is also refuted by adding another “a” to the remaining “c”. The previous model does not prevent an “a” that is also a “c” so long as it is not a “b”. The possibility of an “a” that is also a “c” refutes the “Some Cs are not As” conclusion.

[a [b]]  
 [a [b]]  
 a        [c]  
 a        [c]  
 ...

The temptation at this point would be to conclude that the model does not support any valid conclusion at all. However the conclusion does already appear in the model. The exhaustive representation of the “a”s with respect to the “b”s combined with the mutually exclusive “b” and “c” relation, means there are certainly, “a”s that are not “c”s.

Some As are not Cs

Few people find the notation as helpful as the diagram systems although the notation mirrors the diagram systems exactly.

### *Representational Conventions*

The Johnson-Laird and Byrne (1991) notation is quite different from any of the other examples reported in this project. It relies less obviously on spatial relationships such as containment. Alignment is more important in these tableaux than containment and the borders of the cells are effectively made invisible this way. Possible individuals are aligned horizontally and predicates that can be associated with the individuals are aligned vertically. Square brackets surrounding pairs of terms shows exhaustive representation where there can be no other individuals with the same predicate.

### *Algorithm for solution*

The algorithm for this notation is not greatly supported by the spatial layout of the notation or any other feature of the notation. The learner is left to figure out individuals which fit into the columns of predicates which are still consistent with the premises:

- 1) *an initial model is constructed from the premises which involves noting the possible individuals for each premise and surrounding these individuals with brackets where they have been exhaustively represented;*
- 2) *the search for a conclusion begins by taking a possible solution and looking for possible counter-examples. If no counter-examples are found the conclusion is valid;*
- 3) *if counter-examples are found these other individuals are added to the model until no more are possible. The restriction on what individuals can be added is based on their consistency with the premises, and finally,*
- 4) *the conclusion becomes the relation between the A and C terms which remains after a complete model is constructed or that which was left without individuals which contradicted it.*



The algorithm is simple but the system does not support the learner with useful constraints as the circle systems do. It would be possible to add those constraints, either nominally in the form of extra rules, or if the tableaux were implemented as software, the constraints could be integrated with the functionality of the software. A constraint in Edinburgh-Euler is that when a 'Some As are B' statement has been made, it will not be possible to pull the A and B circles apart. There will always be the smallest area of shading present in the intersection between the two circles. There are no other constraints in Edinburgh-Euler for that premise, so that the A circle could be contained within B, or B within A. In the Johnson-Laird tableaux, the equivalent constraints would be that a line must always exist which has an A and a B label. Others could be added that enclose all A labels, indicating that, all As are Bs, or that enclose all B labels, showing that all Bs are As. One of the main differences between Johnson-Laird's tableaux and the circle system is the method by which constraints are shown. Constraints in all the circle systems can be seen as relatively concrete boundaries of movement. In the tableau, the constraints are the added rules in each line. The tableaux do not rely on a dynamic semantics as the circle systems often do.

### 3.3.5 The predicate calculus

Predicate calculus is a logical language that was formalized in the last half of the nineteenth century by Frege and expanded in this century by Quine, Tarski and Carnap. It has been used widely in cognitive science for representing most aspects of the subject. The predicate calculus is capable of expressing formalizations from many different domains including, linguistics, mathematics, planning and problem solving. The language often seems like a natural way to express certain notions. This expression corresponds to our intuitive understanding of the domain. However it is quite unlikely that anything like the predicate calculus is at the heart of information processing in humans. The inference mechanisms involved in processing statements in the calculus are often not much like the processes humans experience of their own thinking. From the early stages Johnson Laird's research program there is convincing evidence of this (Johnson-Laird, 1983; 1988). Neither are the representations likely to be much like the calculus either. Humans often think and solve problems with representations that directly model the situations they describe (Sloman, 1971, 1993). Diagrams are an example of this kind of representation.

The language is nevertheless, the *de facto* standard of logical languages and is suitably qualified to demonstrate the syllogism. The demonstration again uses the same example as above, All Bs are As, Some Bs are Cs  $\rightarrow$  Some Cs are As.

<u>Problem shown in predicate calculus</u>	
1. $\forall x (B(x) \rightarrow A(x))$	Premise 1.
2. $\exists z (C(z) \wedge B(z))$	Premise 2.
<u>Stage 2. Instantiation of premises</u>	
3. $B_j \rightarrow A_j$	Universal instantiation of (1).
4. $C_j \wedge B_j$	Existential instantiation of (2).
<u>Stage 3. Propositional reasoning</u>	
5. $C_j$	Conjunction elimination from (4).
6. $B_j$	Conjunction elimination from (4).
7. $A_j$	Modus ponens with (6,3).
8. $A_j \wedge C_j$	Conjunction of (5,7).
<u>Stage 4. Generalisation of conclusion</u>	
9. $\exists x (A(x) \wedge C(x))$	Existential generalisation

Figure 9: Predicate calculus method.

Line (1) says that “all builders are architects”, just like all the other examples above. In the predicate calculus this is done with a variable, say “x” that is referred to by predicates, such as bricklayers and architects, by quantifiers like *some* and *all*, and by relations between predicates. The “ $\rightarrow$ ” relation is called material implication and refers to a consistent relationship between the two terms. In other words, the statement “ $a \rightarrow b$ ” is the same as, IF a THEN b. Line (1) says that any “x” that is an architect is also a builder.

Line (2) says that there is at least one canoeist that is also a bricklayer. The “ $\exists$ ” symbol says that there is at least one of the variables that follows. In this case the variable is called “z”. The “ $\wedge$ ” symbol is a conjunction between the two adjoining terms that can be read as “and” as in “A and B”. The whole line could be read as “There is at least one canoeist that is also a bricklayer”. This conforms to the minimum interpretation of the syllogism premise “Some canoeists are bricklayers”.

Lines (3) and (4) simply remove the quantifiers, leaving propositions that can be treated as implied relations between any individuals from the first two lines. Line (3) says that any individual that is a bricklayer will also be an architect, and line (4) says that there is a canoeist that is also a bricklayer.

Lines (5, 6) are the result of applying a rule to line (4) that says a conjunction of terms can be eliminated and the two terms will still be true. The “and” (or conjunction) is removed leaving someone who is a carpenter and someone who is a bricklayer.

Line (7) uses the implication in line (3) “IF b, THEN a” together with the information in line (6) that there is a bricklayer, to get an individual that is an architect. This individual can be combined with the individual in line (5) to form the statement in line (8).

Line (8) says that there is an architect who is also a canoeist, and as we saw from line (2), this is sufficient information to reach the conclusion that, “Some architects are canoeists.”



*Representational Conventions*

The calculus is a rich formalism with many conventions. Not only is the language large but there are also different dialects of the same language. Symbols include: conjunctions, disjunctions, parentheses, negations, conditionals, biconditionals, universal and existential quantifiers. Rules for transformation include conjunction-elimination and disjunctive-introduction. Many simple rules are carried over from propositional logic such as modus-ponens and modus-tollens. The language is large and the applications larger. In manipulations over the short syllogism forms, the required set of conventions is relatively small. Very little of the language is required for solving the syllogism. There are no restrictions on the length of strings or the embeddedness of clauses, in the calculus. This means that the complexity of the language is potentially infinite. The syllogism uses four forms that are quickly written in the predicate language. Sentences are created from five smaller units.

- $\exists x$       There is at least one “x”, or, “x” exists.
- $\forall x$       For all, or, for each “x”
- $A \rightarrow B$     IF A, THEN B.
- $A \wedge B$     A AND B.
- $A \vee B$     A OR B.

Table 3: Predicate calculus equivalents for premises.

Replacing terms with A & B	Predicate calculus equivalent
All A are B	$\forall x A(x) \rightarrow B(x)$
Some A are B	$\exists x A(x) \wedge B(x)$
No A are B	$\forall x A(x) \rightarrow \neg B(x)$
Some A are not B	$\exists x A(x) \wedge \neg B(x)$

*Algorithm for solution*

Several ways are available to solve the syllogism with the predicate calculus. Syllogisms can be solved in the calculus in the same way that all inference can be made either by contradiction, that is by proving that the negation of a conclusion is not consistent with the premises, or by showing that a statement is possible from the premises.

- 1) *Distinctive premises are formed as shown in the table above for which there are essentially only four alternatives at this point;*
- 2) *instantiation of premises is allowable in order for propositional reasoning to occur;*
- 3) *propositional reasoning involves the use of several standard rules, and*
- 4) *conclusions are generalized either to existential or universal quantification.*

Although there are only four stages here this does not represent a direct comparison of complexity with the other systems. Each propositional step has rules that could be shown separately, increasing the detail of the process and increasing the number of steps involved. This system is possibly the most complex of all alternatives, although we do not have a good characterization for comparing it with the others, and so cannot easily say, for example, how much more expressive it is compared with TARSKI'S WORLD or any of the others.

### 3.3.6 Tarski's World

TARSKI'S WORLD is an interactive software program developed by John Barwise and John Etchemendy to be used with the book, *The Language of First Order Logic* (Barwise & Etchemendy, 1990). Unlike the other systems in this chapter, it was designed specifically to be used as an interactive software application. For the symmetry of this chapter however, the following pages describe the system in the abstract. A fuller description of the functionality of the program is provided in Chapter 5. The system consists of two formalisms, a blocks world and sentences written with the first order calculus with the equality relation. Premises of a syllogism must be translated into sentences and worlds that can be understood by TARSKI'S WORLD. The example syllogism (as above), is used here to illustrate. The problem is, "All Bs are As and Some Bs are Cs  $\rightarrow$  Some Cs are As." The process begins by translating the premises into usable terms that make sense to the program. The first premise can be translated into the statement:

$$\forall x (\text{Cube}(x) \rightarrow \text{Large}(x))$$

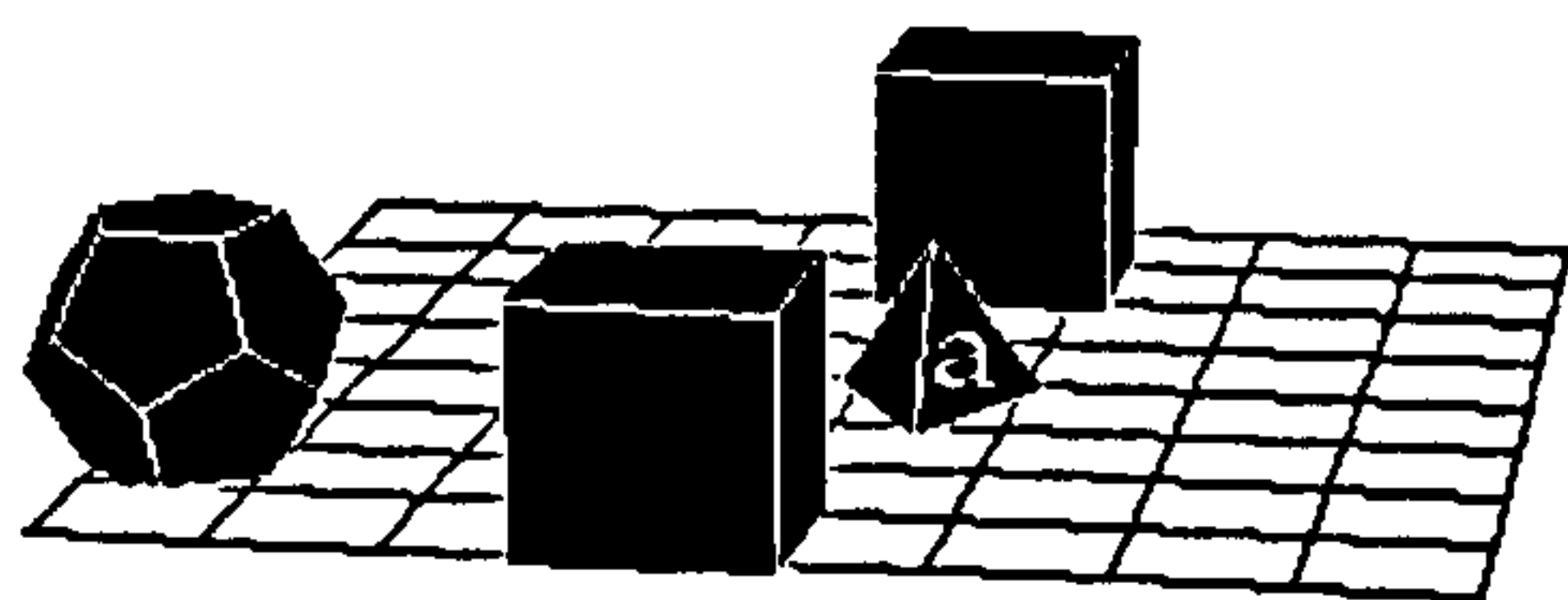
In choosing this representation, we have translated the B term into a cube and the A term into Large things. The sentence reads "All cubes are large". The second premise can be shown where  $\text{Backof}(x,a)$  means x is behind a.

$$\exists x (\text{Cube}(x) \wedge \text{Backof}(x,a))$$

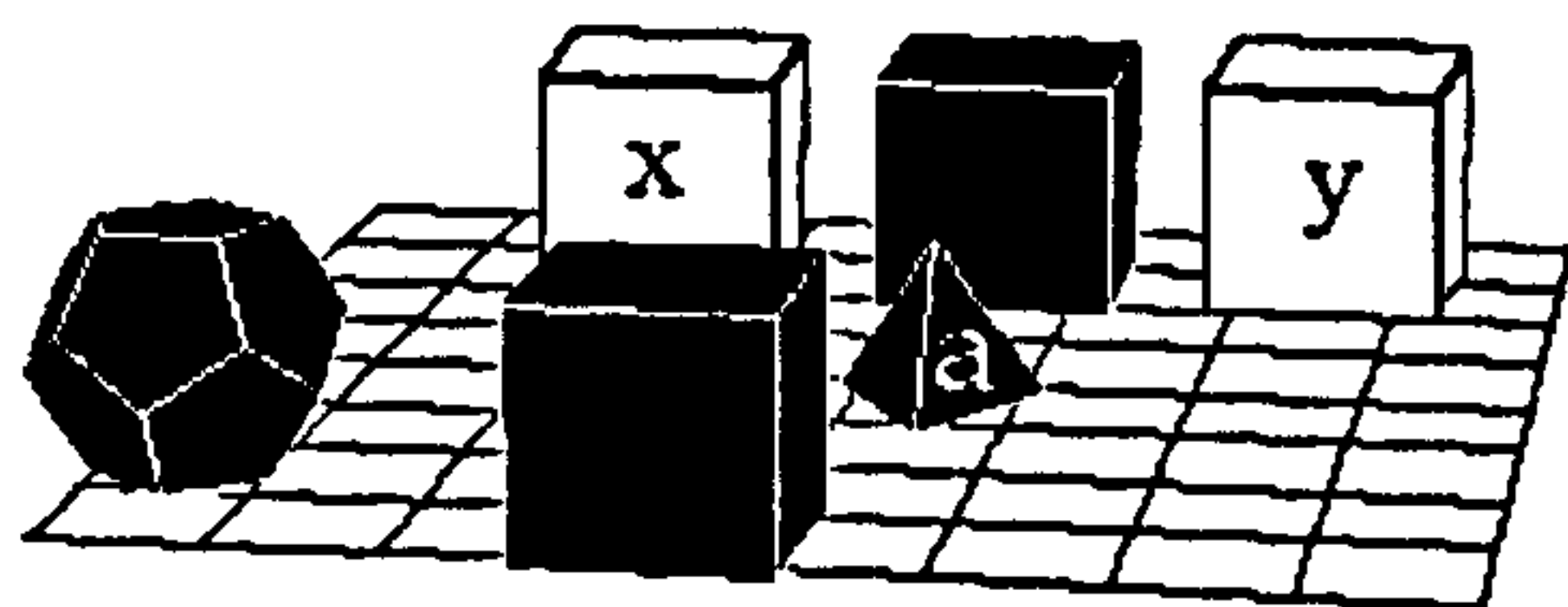
In full, the sentence should read, "there is a cube which is behind 'a'". Both premises have been translated into sentences that can be understood by TARSKI'S WORLD. The next stage is to create a diagram, micro-world, or situation in which these sentences can be true. We determine the validity of a conclusion by testing it in all situations where premises are true. If the conclusion remains true in all such situations it is valid. The trick in using TARSKI'S WORLD is to make the premises true and find a conclusion which is also *persistently* true. The learners must change the diagram in as many relevant ways as possible, trying to falsify the conclusion without altering the premises. If the learner fails to falsify the conclusion under



these conditions, then the conclusion must be valid. Here is a world where the premises are true.



In this situation the premises: (a) All Cubes are Large, and (b) Some Cubes are behind 'a' are true, and the conclusion, Some things are behind 'a', cannot be made false whilst those premises remain true. By adding more cubes behind 'a', the truth of the first premise will not be altered. Neither will the truth or the validity of the conclusion. Several changes to the world would violate the truth of the premises. A world with only a small cube, a world with no cubes, or a world with no cubes behind 'a', would all violate one or the other of the premises and would not therefore be a good representation of the syllogism.



All cubes in the first world are large and there is a cube which is behind the Tetrahedron marked 'a'. Whatever way the picture is changed keeping the premises true, there will always be a large cube behind 'a'. Several large cubes could for example be added such as 'x' and 'y' above and several more tetrahedra. The Dodecahedron could be removed. Whatever we do to the diagram whilst the premises remain true, we cannot alter the truth of the conclusion that "Some Large things are behind 'a'".

Unfortunately TARSKI'S WORLD allows for what are called vacuously true statements. This is a problem for the syllogism as Aristotelian logic does not allow such inference because it assumes no empty sets. The problem in using TARSKI'S WORLD is that the statement All Cubes are Large will be true even whilst there are no cubes in the world. In order for TARSKI'S WORLD to work the provision is made that whenever a statement is positively made about an individual, at least one example of that individual must exist in the world created to test the syllogism.

### Representational conventions

In TARSKI'S WORLD we have a set of conventions based on the predicate calculus supplemented with a situation which can be related to any sentence in the calculus. These conventions then follow the spirit of heterogeneous logic developed by Barwise and Etchemendy (1995). The iconic representation of tokens of Tetrahedra, Cubes, and Dodecahedra can be in three different sizes: small, medium, and large. Worlds something like

a chess board exists where eight by eight positions are available for objects. Relations between objects are shown using a restricted set of relations: (a) BackOf(a,b), (b) FrontOf(a,b), (c) LeftOf(a,b), (d) RightOf(a,b), (e) Smaller(a,b), (f) Larger(a,b), and (g) Between(a,b,c). These representations allow the construction of premises and conclusions for a syllogism: (a)  $\exists x \text{ Cube}(x) \wedge \text{Large}(x)$ , (b)  $\forall x (\text{Large}(x) \rightarrow \text{BackOf}(x,a))$  and (c)  $\exists x (\text{BackOf}(x,a) \wedge \text{Cube}(x))$ .

#### *Algorithm for solution*

To solve a syllogism using this representation we use a form of proof by cases. There are several steps in using TARSKI'S WORLD and several additional rules which make the process even harder. These extra steps are partly because of the situated content of the world being used and partly because of the logic which is assumed by TARSKI'S WORLD.

- 1) *form representative premises using predicates and relations available in TARSKI'S WORLD.*
  - *Relations with transitivity, asymmetry and reflexivity should be used only when certain that there is no interference between predicates;*
  - *A premise indicating existence of an individual requires an individual to be present in the world, even though TARSKI'S WORLD will evaluate as true, a sentence in a world without such an individual;*
  - *Representing predicates must have the same type relationship as referent predicates for example, the premise, "All Cubes are Tetrahedra" is a well formed premise according to the rules of the syllogism but is impossible to represent in the TARSKI'S WORLD;*
- 2) *when appropriate premises have been created a conclusion is postulated. Premises must be well formed sentences and true, and*
- 3) *each of the four possible forms for the syllogism are postulated and*
- 4) *alter the world by adding new objects and moving objects on the board. Each new situation must maintain the truth of the premises. Each time, try to make false the putative conclusion.*
- 5) *If we cannot make false this putative conclusion, it is then the valid solution.*

This algorithm is complicated mainly by the translation process where terms in ordinary language must be translated into the terms and relations available in TARSKI'S WORLD. The restrictions on which terms can be used make this system a difficult one to use. The requirement to know the limited subset of the notation will put many people off using it. The process of proving by cases makes sense and the situation based approach here is likely to prove attractive to learners.

#### *3.3.7 Carroll's 'Game of Logic'*

The last system was developed by Carroll (1958) and works with four quadrants in a grid that divide the square across the horizontal and vertical axis. The upper half of the grid is a place holder for members of the A term group and the lower half for individuals not members of the A group. The grid is divided vertically and a place holder provided for

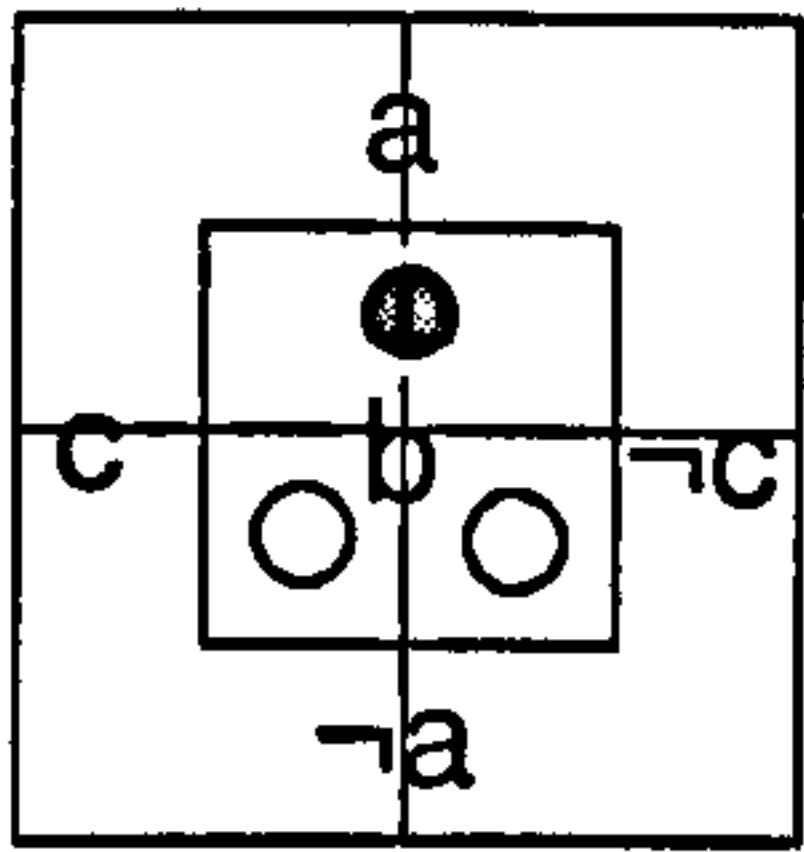


members of the C group to the left and for those not in C group to the right. In the centre of the grid a square contains all members of the B term group. All regions around this inner square are considered not members of the B term group and so all possible relations between the three terms are possible in some region of the grid.

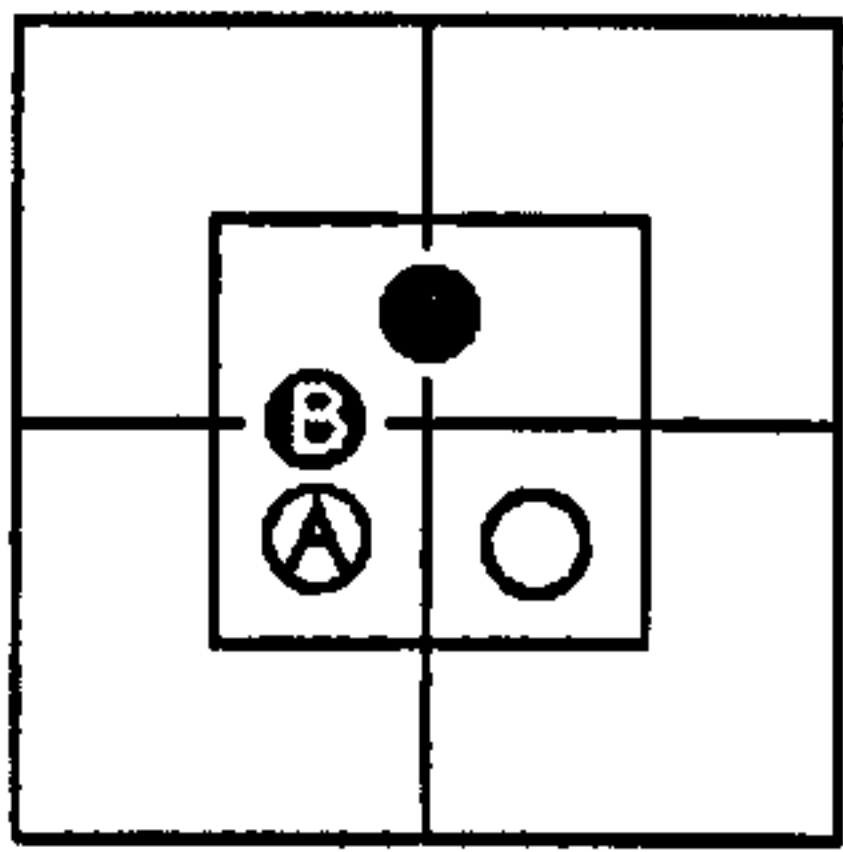
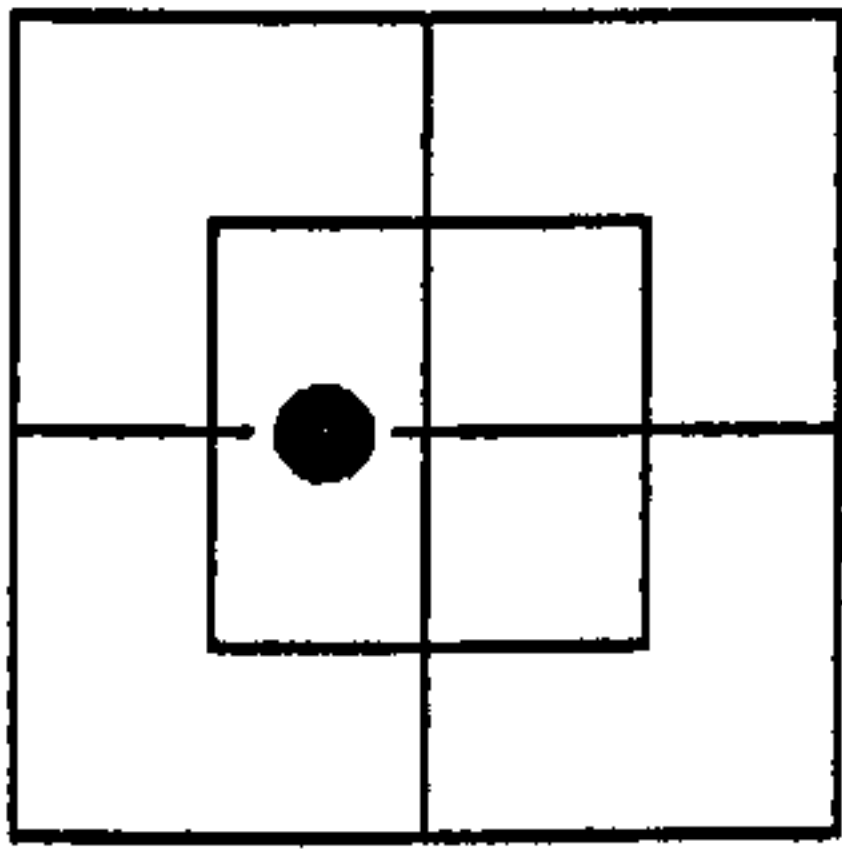
We use Carroll’s system to solve syllogistic problems by placing counters in the regions and on arcs adjoining regions. In the problem used above, All Bs are As, Some B are C → Some A are Cs, Carroll’s system looks like Figure 10.

*P1. All Bs are A. The premise implies that there are no Bs that are not As, hence the unshaded counters.*

*The combined diagram shows counters from first two premises.*



*P2. Some Bs are Cs. The premise allows that the B may be either A or not A.*



*C. Since the A counter describes a definitely empty region, and B is either in that region or immediately above, it must be in the region above.*

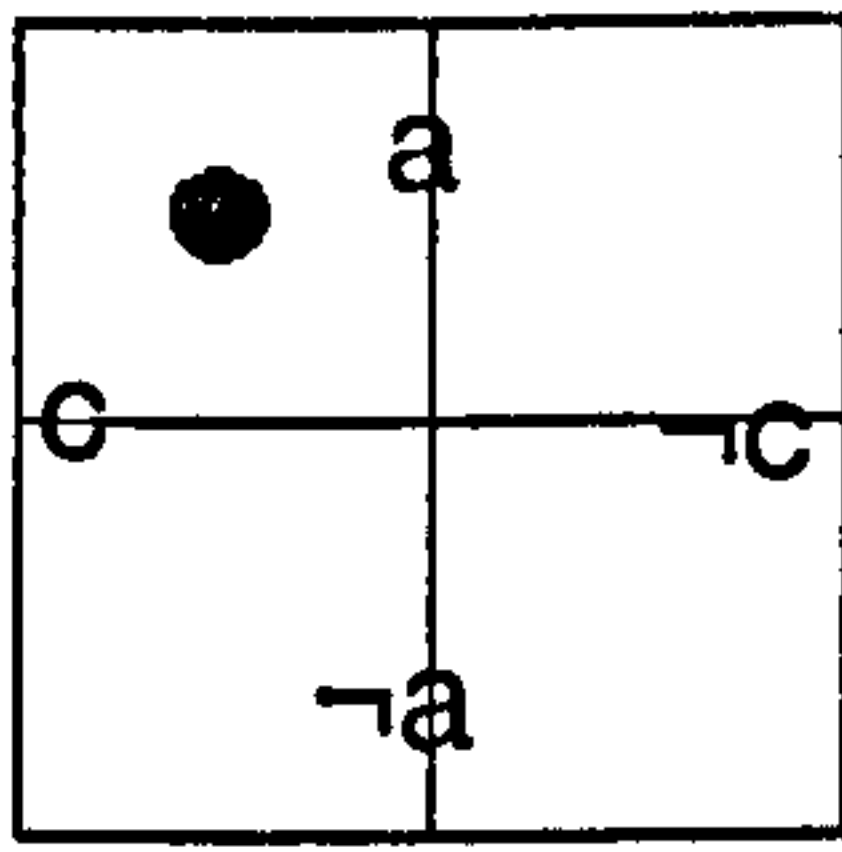


Figure 10: The Carroll method.

Representational conventions

The two essential conventions used in Carroll are the location and shading of the counters. The grid is divided so that counters placed in the upper half of the grid refer to individuals that are “a”s. Counters in the lower half refer to individuals that are not “a”s. Counters to the left refer to “c”s and to the right refer to those that are not “c”s. The central square is used to illustrate individuals that are “b”s and anything outside of the central square refers to individuals that are not “b”s.

Placing a shaded counter within a region indicates the known existence of an individual with the properties associated with that region. An unshaded counter within a region demonstrates a region that is known not be inhabited. The combination of counters in regions and on lines separating regions, can demonstrate any of the premise types and any other relation between three terms.

The first premise above, “All Bs are As,” uses three counters, one shaded the other unshaded. The shaded counter shows there are known “b”s that are also “a”s (the existential claim). The shaded counter is in the “b” and “a” region, but is located on the

line that separates “c”s from not “c”s. The two unshaded counters show the stronger (universal) claim, that there are no “b”s that are not “a”s.

The second premise shows “Some Bs are Cs”. The shaded counter indicates a known individual that is “b” (the shaded counter is shown inside the “b” inner square). This known individual is also known to be a “c” (the shaded counter is not only in the “b” square, but also in the affirmative “c” side of the grid).

Combining the two premises together, yields the third diagram in the figure above. The two labeled counters in this picture illustrate a contradiction that has to be resolved for the conclusion to properly fall out. The shaded counter (labeled B) indicates that there is an individual in one or other of the immediate neighbour regions. The unshaded counter (labeled A), indicates that there is certainly no inhabitants in one of those neighbour regions. The diagram is resolved by shifting the shaded counter (labeled B) into the upper neighbouring region.

The conclusion between the A and C terms, then falls out, Some As are Cs. The final diagram shows just the relation between “a”s and “c”s.

#### *Algorithm for solution*

The process of solving the syllogism with the Carroll system can be characterized as a four stage algorithm:

- 1) *form representative diagrams for each of the premises for which there are four patterns which can be in different orientation depending on the mood of the premise;*
- 2) *register diagrams by adding tokens from each premise diagram into one;*
- 3) *resolve regions where there is ambiguity and confirmed individuals for example where there is a shaded counter separating two regions and one of those regions has an empty counter in it then move the shaded counter into the non-empty region.*
- 4) *the conclusion is read from the diagram by determining any conclusion that fits the patterns of the premises but between the A and the C term.*

Patterns of counters found in the conclusions of Carroll, are the same as patterns of premises. A study reported later in Chapter 6 shows that one of the most pervasive errors in using the Venn system is to add the representation for the conclusion to the diagram as if it were another premise. In the Venn system the conclusion is read from the diagram but does not use the same representation as used to represent the premises. This is an advantage of Carroll over Venn. An error found widely in Venn is not possible in Carroll. If the student tried to represent an invalid conclusion with Carroll they would see that the valid one was already present.



### *3.4 Implications for study.*

The brief treatment of syllogistic reasoning in the early part of the chapter, helps to convey the components that constitute this kind of reasoning and to see the value in providing instructional tools that improve human performance. The account of explanations for human failure establishes a preliminary basis to support teaching and learning. Theories that compete to explain the root of several errors, contribute to improved teaching by supporting better learner diagnosis.

Treatments of seven strict systems for demonstrating the syllogism provide a detailed characterization of each. The descriptions indirectly illustrate several ways that these systems are different from each other. These variables are each likely to affect the benefit of the systems for learners:

- 1) the distribution of graphical and sentential representation used,
- 2) the scope of expression possible,
- 3) the distribution of dynamic, versus static conventions used,
- 4) the complexity of the algorithm for solving the syllogism,
- 5) the situatedness of the representation,
- 6) the ease in understanding of the conventions and operations used,
- 7) the divergence of conventions used to read premises and write conclusions and,
- 8) the deviation of the underlying logic,

The first five of these emergent variables reinforce the selection of the computational explanation of graphical reasoning presented in Chapter 1. These factors are very closely related to the specificity principle. The scope of expression for each system (factor 2) is, according to the specificity principle, a direct result of the graphical conventions used in the system (factor 1). Graphical representations are supposed to restrict the possible expression of a system and result in systems that are much less able to express a variety of material than systems with more abstract conventions. Up until this time there has been no method available to accurately measure either of these factors. The procedure developed in Chapter 4 is equipped to characterize this second factor, albeit with a specialized and limited set of representational systems.

The distribution of dynamic and static conventions (3), may affect the value of the specificity principle. Modifications to older systems that make the graphical lexicon somewhat more simple, also tend to make the syntax more complex. This suggests a trade-off between the simplicity of the dynamic and static conventions. This is illustrated by the

differences between the older and more recent Euler systems. The newer system is characterized as simpler because there is only one diagram for each premise type. In the older system a premise type was shown with several diagrams that all illustrated possible worlds that could be true for that premise. Solving a problem could amount to dealing with 18 or more diagrams (*see* 3.3.2, this chapter). Assessing the two systems using the specificity principle leads to the conclusion that the newer system is simpler because it is only able to represent the statements of the syllogism and not the statements that refer to all those 18 possible world diagrams.

However the distribution between dynamic and static conventions (3) has been altered in this transition to the new Euler system. The processes used to manipulate the new Euler system diagrams involves a complex set of comparisons and restrictions of circle movement (4). The following chapter describes a method for applying the computational account of diagram processing to evaluate real diagram systems. The method does not, however, address this issue. The section on further work presented in Chapter 8 identifies a field of mathematics that deals with the measurement of complexity in algorithms (*e.g.*, Bovet & Crescenzi, 1994). A more complete treatment of diagram systems will require a treatment of algorithm complexity.

More situated representations (5) have a tendency to be graphical and hence (according to the theory) more specific. Situated representations are intended to be authentic and similar in form to the represented object, or to use concrete objects to show abstract material. The cubes and tetrahedra of are situated because they use graphical representations of tangible objects that are necessarily graphical. The analysis in Chapter 2 described the term 'homomorphic,' roughly as the similarity of form between representation and referent. Each of the diagram-based systems described in this chapter come close to the criteria outlined there. The mappings between diagram elements and referents preserve the relations in the domain, producing rigorous representations. However, Venn and Carroll are both able to illustrate situations where there is ambiguity. In Venn, the star or cross notation on the boundary of two regions allows for the actual individual to be in one or the other region (an empirically impossible situation). The idea of an homomorphic representation is similar to the account of situated representations. Homomorphic representations are very likely to be graphical and hence subject to the same specificity principle outlined already.

These factors (from 1 to 5), are plainly interconnected, and have some bearing on the value that the specificity principle may have on selecting useful diagrams for learners. The remaining factors (6 through 8) are unrelated to the specificity principle. We notice in the Venn system that reading conclusions from a diagram requires a different set of interpretations than those needed to annotate the diagram with a premise (7). This may be alarming to a learner because the forms of the premise and the conclusion are exactly similar.



At minimum we would expect an increased lead time to learn the system. Indeed, the same problem does not exist in any of the other systems.

The final factor appraises differences between the systems that result from inconsistencies in the underlying logic (8). Some of these differences can be controlled in empirical study and systems can be modified to reduce the differences. This problem is most confounding when systems have been implemented in software and cannot be easily changed. The main area of divergence is in the acceptance of the implication from All to Some. This issue is sometimes known as 'the problem of existential import' and is also described as a choice to consider the domain of interpretation as empty or non-empty. The Venn system described in this chapter chooses the traditional Aristotelian approach which assumes non-empty domains, allowing the implication from All to Some. The software used to demonstrate Venn in Chapter 5 (Venn, 1991), chooses not to assume non-empty domains. This means that even when premises include universal statements like 'All diagrams are formal' it is not possible to assume that any diagrams exist. The TARSKI'S WORLD design is similar. Statements like 'All cubes are large' are taken to be true in TARSKI'S WORLD, even when there are no cubes in those worlds.

These emergent factors, particularly the first five outlined, have reinforced the choice of 'specificity' as a way to explain the benefit of graphical systems. Investigation of this principle should yield suitable guidelines and rules for micro-media selection.

### *3.5 Methodology.*

The method of investigation is motivated by experimental but pragmatic opportunism. Decisions about design and data to collect are driven by a general goal to provide better understanding in a small field of technology-based learning. The precise hypotheses of empirical designs sometimes take lower priority than new and unexpected results that emerge during data collection. Sometimes these peripheral results direct subsequent stages in the empirical designs. Although early hypotheses are tested and results reported, several other valuable findings emerge during the process.

Different disciplines have contributed methodological standards to those used in the project. Research in learning and instruction has often used detailed qualitative approaches to data collection. The methods and principles of 'phenomenographic' study illustrate the benefit from involving learners as a rich source of data (Marton & Saljö, 1984; Entwistle & Ramsden, 1983; Ramsden, 1988; Laurillard, 1991). Phenomenography has even directly contributed to better understanding of learning with representations (Lindström, 1980). In Lindström's work, small groups discuss specially designed and focused questions, and the dialogue is analyzed by iterative search and peer review. These studies usually result in 'outcome spaces' or small collections of misconceptions that are widely found in the population studied. There is considerable discussion about the methods used in this research



framework (Bowden & Walsh, 1994) and the project was not intended to add to that debate. The project does however use similar procedures to those reported in phenomenographic studies but with the primary intention of providing detailed information that can be used to diagnose student errors in syllogistic reasoning.

The subject of analysis is different in these studies from those generally found in the phenomenographic program. A pilot study reported in Chapter 5 uses transcriptions of discussions with learners talking about syllogistic reasoning. However the data set is very small and the discussion is more instructional than interrogative. In this project, most of the qualitative data that leads to a form of 'outcome space', comes from diagrams produced by learners while solving syllogistic reasoning problems. The other main form of data collected is also visual. Approximately 8 hours of video footage recorded the activities of selected learners developing skills in syllogistic reasoning, sometimes while using the software, and sometimes without. This data was analyzed with a qualitative analysis tool that focuses on the time based properties of the data set.

Research in learning and instruction has sometimes used other qualitative methods like card-sorting (Gilbert, Watts, & Osborne, 1985). This project adopts a card-sorting approach and with some modification to the older method, has added to the scope of results that are now possible with card sorting studies.

The methods of experimental psychology can also be found in Chapters 6 and 7. These studies begin by stating an hypothesis. Groups of similar and appropriately matched individuals are treated to different learning experiences. Differences in the outcomes from those treatments are measured for significant variation and the differences are attributed to the dependent variables involved.

The methodology of the project is not only empirical. Taking a theory of graphical information processing and applying it so that decisions can be made between competing systems, requires a certain amount of interpretation and analysis. The theory of information processing is stated in the literature so that it reads quite rigorously and appears to provide a ready set of rules that could easily be developed into formula. The descriptions of systems above has already shown that interpreting the language used in these definitions can be difficult. When the meaning of the principle has been decisively settled, its application to graphical systems, is still problematic. The next chapter shows how a "wire-frame cube" helped to apply the metric as a kind of scaffolding or 'intermediate representation'.

The inductive process of interpreting statements made by information processing models is an on-going developmental one. The project makes a few small steps in this evolution. The first step was to describe the 'specificity' principle in the context of other candidate explanations for diagram benefit. The second stage comes in the following chapter. Chapter 4 develops a method of interpreting the specificity principle as a formula that can be



applied to certain kinds of graphical systems. The remaining step is in the reflection and suggestions for further work given in the final chapter.

## 4 Specificity Applied

### *4.1 Quantifying specificity.*

Communication in an instructional environment involves the use of languages to illustrate the material to the learner and for the learner to demonstrate comprehension. Chapter 2 described the role of graphical representations as information presentations, and as tools for human processing. Three approaches to information processing provide insight into diagram processing. One approach claims there are identifiable features of representations that lead them to be pedagogically useful. This specificity principle is appealing and convincing. It suggests an accurate and formal interpretation that could be applied to different systems. The account does not include any description for applying it to graphical systems and so the process described here relies on an interpretation of the principle.

The results of treating different systems in this way will inform choices between alternative systems. This information constitutes guidelines for instructional systems development and precise rules for automating media selection in instructional systems. As a part of the instructional system these rules form the beginning of intelligent multimedia (Maybury, in press), and modality choice decision support systems for instructional design (Dijkstra, 1997).

The work in this chapter demonstrates the meaning of the specificity principle, interprets the principle as a simple mathematical formula, applies it to three graphical systems for showing the syllogism, and reports the results. In doing these things, the chapter provides the conditions for testing the specificity principle in an empirical framework.

The idea of 'specificity' originated with the SIGNAL project (Stenning & Tobin, 1993; Stenning & Oberlander, 1994). Their work outlined a theory of graphical representation in reasoning. Roughly speaking the theory claims that representations that are capable of limited expression are the most useful, as long as they are expressive enough to represent the material communicated.

Graphical representations are one sort of representation that exhibit specificity. Graphical systems integrate small numbers of lexical items that can only be combined and manipulated with a limited set of operators. In this way the graphical systems 'compel' the allocation of information into simple and uncluttered representations that are easier to understand than other systems that allow more abstraction. Specificity is a feature of graphical and indeed, all representations, that influences the ease with which they can be processed. Since processing information is a prerequisite for learning (although clearly not



the same), it is likely that specificity should have a positive affect on the choice between alternative graphical systems used for learning. Stenning and Oberlander (1994) believe it is 'specificity' that helps to explain the didactic effectiveness of graphical systems used for teaching abstract reasoning.

Applying this principle to three systems that demonstrate syllogistic reasoning, raises several difficulties. The complexity of counting the variety of states in the diagrams leads to problems tracking the counting argument. A wire-frame cube overcomes these problems by supporting identifiably similar parts of the process. The cube is used as an 'intermediate representation' so that the counting is clearer and less ambiguous than without it. There are valuable insights in the reflections on the use of the cube that also apply to other problem solving situations.

Another difficulty in the counting process arises from the artificiality of limits imposed on the size of systems. The Venn diagram system (Chapter 3) invariably has only three circles. However there is no fundamental reason for this limitation. Other presentations of the Venn system sometimes use more than just three circles (*e.g.*, Gardner, 1958). Two separate methods of counting the systems overcome this issue. The first gives an estimate of specificity based on the most used version of the system. For Venn this is a system that uses three circles. The second draws on a restricted version of the normal system that reduces the number of tokens used to the minimum required. This produces a smaller ratio (although not one-to-one).

The Office Indicator System (OIS) is a simple graphical communication system. It illustrates the specificity principle and the method used to calculate the values of each system. The OIS has a small and fixed number of symbols. Communicating information with the system requires only a small number of possible states. The exact number of states possible depends on the exact rules use to define the operation of the system. Three of the systems used to illustrate the syllogism presented in Chapter 3 are later treated with this analysis. These systems are similar to the OIS since they all have a fixed upper limit of system states.

This information provides the grounds to test the first hypothesis. The hypothesis that learning with graphical representations that have a minimum expressive ability will yield the best outcomes.

#### *4.2 The office indicator system.*

The office indicator system is a simple graphical language that illustrates the key ideas of the specificity principle. In an hypothetical scenario, the system is fixed to the external door of an office that is shared by several workers. The graphical lexicon informs visitors about important aspects of the situation inside the office. The exact configuration of the graphical language determines what information can be determined from the sign. Some

configurations allow the visitor to see how many people occupy the space, but not exactly who they are. Other configurations illustrate who is in as well as who is engaged in a meeting.

The representation consists of four quadrants in a square. Each quadrant is either shaded or unshaded. Different systems are built from the rules used to interpret these images. One configuration shades one quadrant to show one office worker. The quadrant shaded is irrespective of which worker is present. The visitor in this case cannot tell from the illustration exactly which of the four workers occupies the office. Another configuration maps the seating position of each worker to the particular quadrant of the graphic. In this system the quadrant belonging to the individual is shaded when they are present. In this system the visitor can tell exactly who occupies the office as well as where they sit.

The office indicator system is a little different from the syllogism systems that appear later in the chapter. These systems are inherently static. There are no rules to constrain possible graphical states. They are not interactive.

Four variations of the office indicator system are described here. Each has a slightly different set of rules and constraints that influence how the information can be interpreted. This affects the possible number of states and hence the expressive ability of the system.

The first variation is called OIS-1. Diagrams that would be the same when rotated by 90°, 180° or 270° are considered to be the same diagram. These diagrams are called ‘rotationally equivalent’. The total number of states in OIS-1 is six. These states are all illustrated in Figure 11.

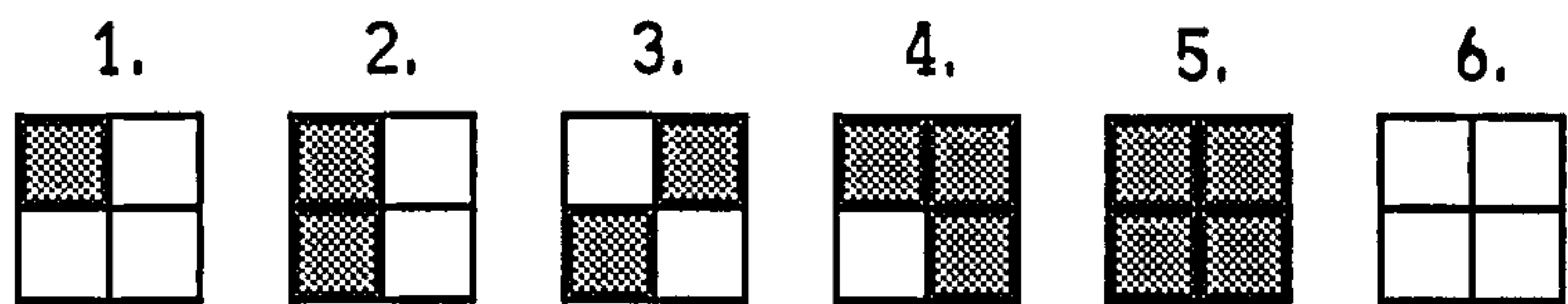


Figure 11: Office indicator system (assuming rotational equivalence).

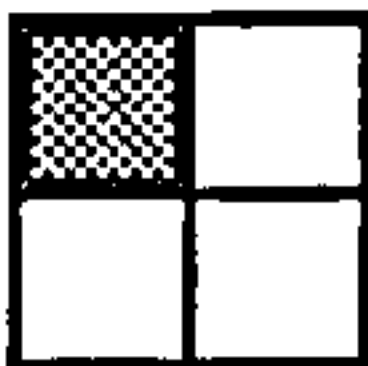
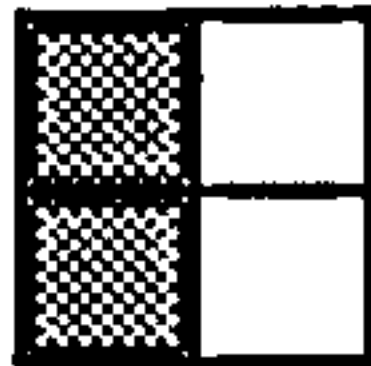
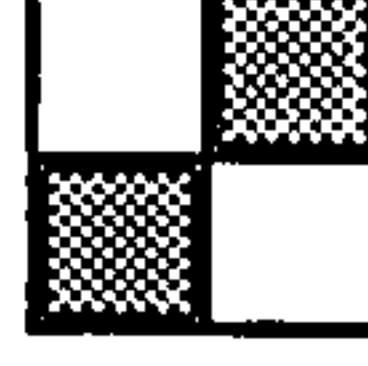
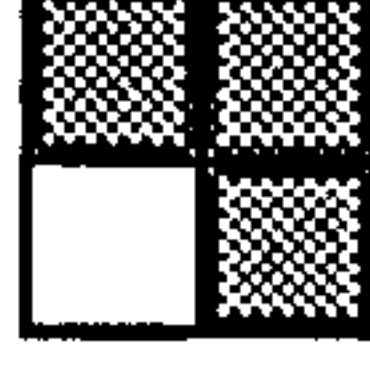
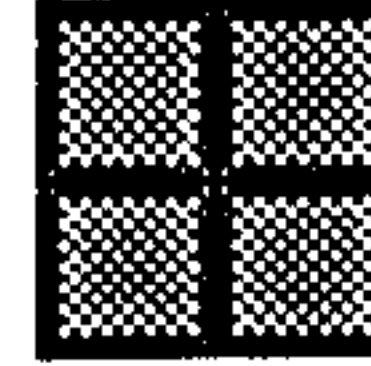
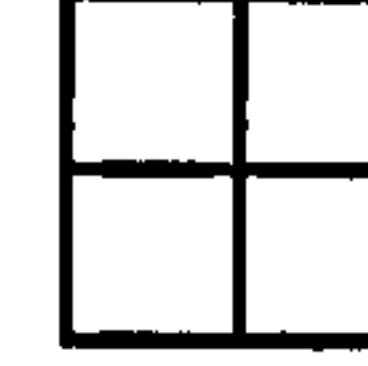
The corresponding situations inside the office include several more states than the six available from OIS-1. The four workers, counted as simply either in or out, sum to  $2^4$  possible alternative situations. Each of the illustrations corresponds to a number of possible situations. Diagram (1) maps to four different states (for each one of the workers being inside). Diagram (2) corresponds to four different situations (for the two possible pairs of workers in adjacent desks). Diagram (3) corresponds to two possible situations (the two pairs of workers in diagonally opposite desks. Diagram (4) corresponds to four situations (the four possible absentees). Diagram (5) and (6) can apply to only one situation each (the office is either full or empty). In these final cases, the illustration is adequate to represent all the information that is required. Measuring the number of states possible and the number of different states the system represents, produces a ratio of 6:16 or 0.375. The description of OIS-1 shows that in



this configuration the system is less than adequate to demonstrate the required information. The system is underspecific. A visitor arriving at the sign with a particular question cannot guarantee a determinate answer from the diagram.

New conventions added to the system make the system more expressive. The rotational equivalence condition was a way to restrict the number of states. When this condition is removed, the number of states are increased. Without rotational equivalence there are 16 different states in the representation and, of course, the same 16 states in the world to represent. This produces a ratio of 16:16 or one-to-one. The diagram perfectly represents the states required and is therefore optimally specific. Visitors to the office in this case can tell who is in the office in each situation.

Table 4: Numbers of states in the office indicator system.

Situations	Alternative diagrams						Totals
							
OIS1&2	4	4	2	4	1	1	16
OIS3&4	$4*2 = 8$	$4*2^2 = 16$	$2*2^2 = 8$	$4*2^3 = 32$	$2^4 = 16$	1	81

The same system becomes less specific by requiring it to communicate more information. The visitor, for example, would benefit from knowing more than just that the individual is inside the office. The visitor will avoid disturbing a meeting or telephone conversation if they are aware of the situation inside the office. The third variation (OIS-3) requires, but is unable to communicate, this new information. The increase of possible combinations of occupant’s privacy status, increases the effective size of the domain. This is the situation shown in OIS-3. There are still four individuals in the office. Removing rotational equivalence increases the available tokens so that there are enough to show all workers in the office. There are not enough tokens to show their ‘privacy status’. Each individual has a privacy status, either open or closed. Instead of the simple 16 states in the previous cases, there are now 81 alternative states in the office world. There is a shortfall between the new information required and the available states in the system. OIS-3 supports 16 states but 81 situations are required. The OIS-3 is now underspecific, yielding a ratio less than one (roughly, 0.2).

Again, adding new conventions to the system effectively adjusts the system specificity. The new representational convention allows more representational states and more communication. Two levels of grey shading in each quadrant are sufficient to show this new

information. This would yield three possible states for any one quadrant, empty, grey or black. These states now represent an unoccupied desk, a desk occupied with a worker in a meeting, and a desk occupied by a worker available for visitors. This final situation is called OIS-4 in the table below. The system now has a one to one ratio of system states relative to required information states. There are 81 states in the system and 81 configurations of the world inside the office. The visitor will now be able to make an informed judgment about several aspects of the situation inside the office.

The four systems illustrate variations of representations and the consequential changes in information that is communicable. The four systems are examples of a related class of similar representations. Each system differs in the scope of allowable values for the tokens they use.

Table 5: Features of office indicator system.

	Expressivity	Domain size	Specificity
OIS <sup>1</sup>	6	16	0.375
OIS <sup>2</sup>	16	16	1.000
OIS <sup>3</sup>	16	81	0.198
OIS <sup>4</sup>	81	81	1.000

*Note:* Specificity is at unit value when the expressivity of the system is the same as the domain size.

Representational systems support communication of different quantities of information. This measure is called expressivity. The comparison of the expressivity of a system with the content of the communication, produces a ratio. The ratio is an interpretation of the specificity principle and has been verified with its author (Stenning, personal communication, August 29, 1995). The “specificity” principle indicates that systems which perform well with the metric should be better for learners than those that perform poorly. The office indicator system has illustrated how the metric can be applied to a system for showing simple information about office inhabitants.

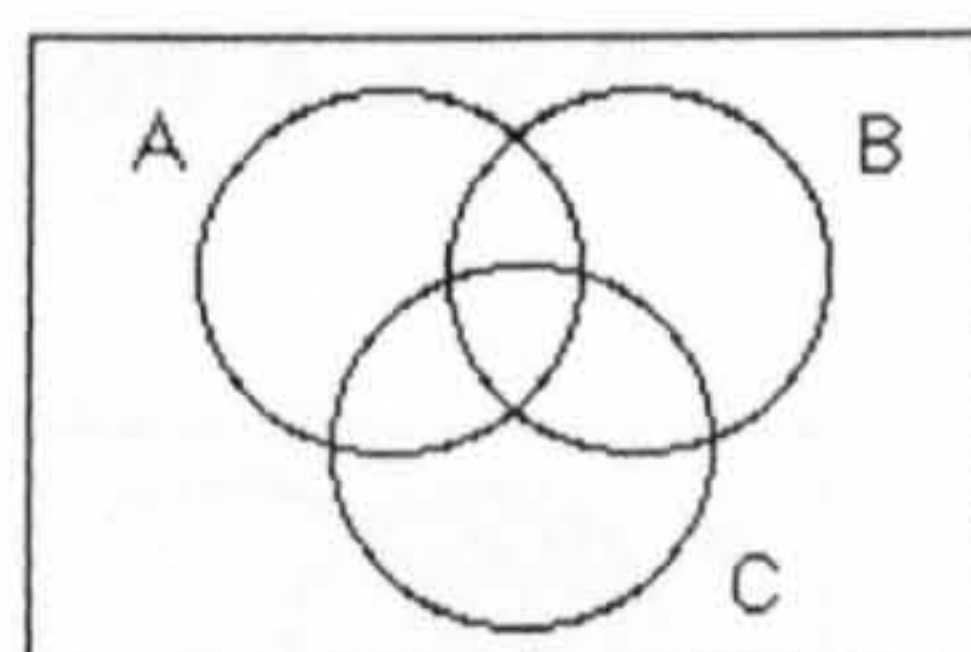
The principle topic of learning for the research was described in Chapter 3. Syllogistic reasoning can be demonstrated with the diagram systems also shown in Chapter 3. The rest of this chapter applies the specificity metric to three of the circle systems for showing the syllogism. The systems referred to here are the same abstract systems discussed in Chapter 3. The software applications that use these representations are described in Chapter 5.

4.3 Analysis of Venn circle system.

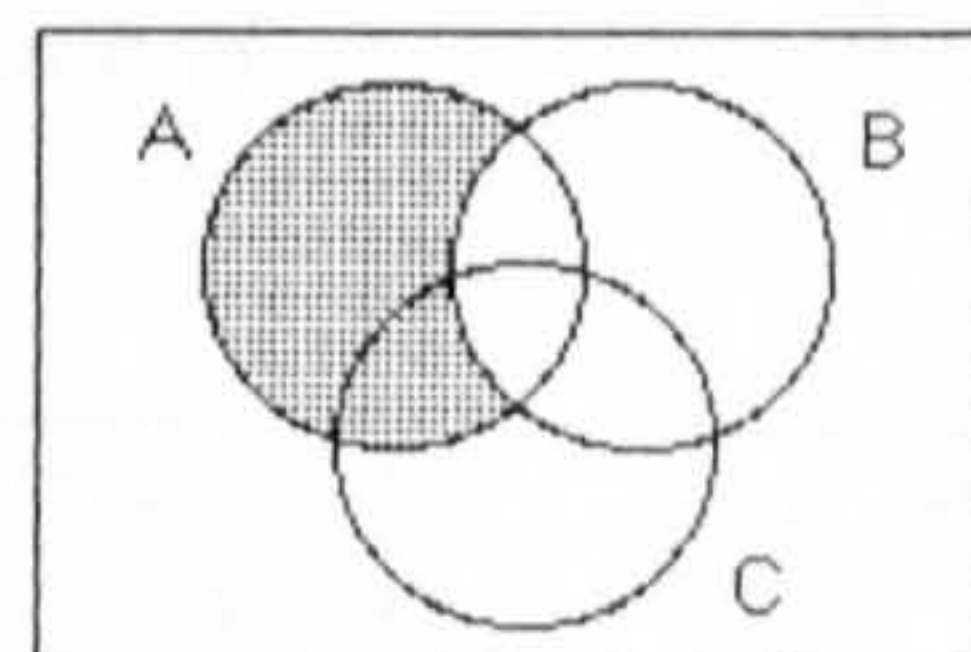
The Venn system starts with this basic diagram below (a). Assertions can be represented on the diagram in one of two ways. Universal assertions are depicted by shading out



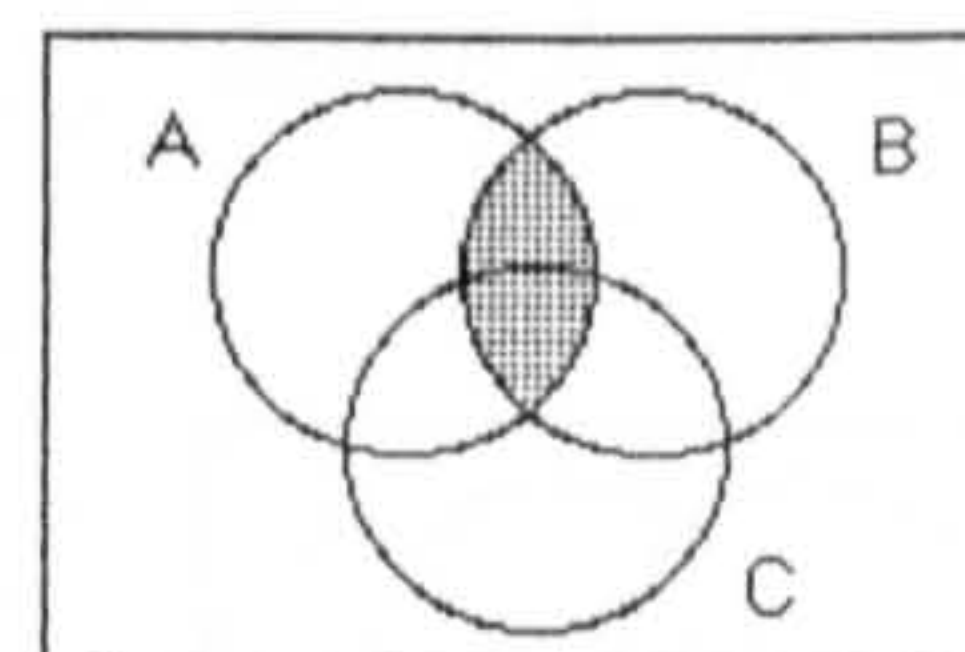
infeasible regions. In diagram (b), shading denotes the fact that All A are B since the part of A which is not also in B is shaded and therefore infeasible. Diagram (c) indicates that No A is B, the intersection being precluded by shading.



(a)

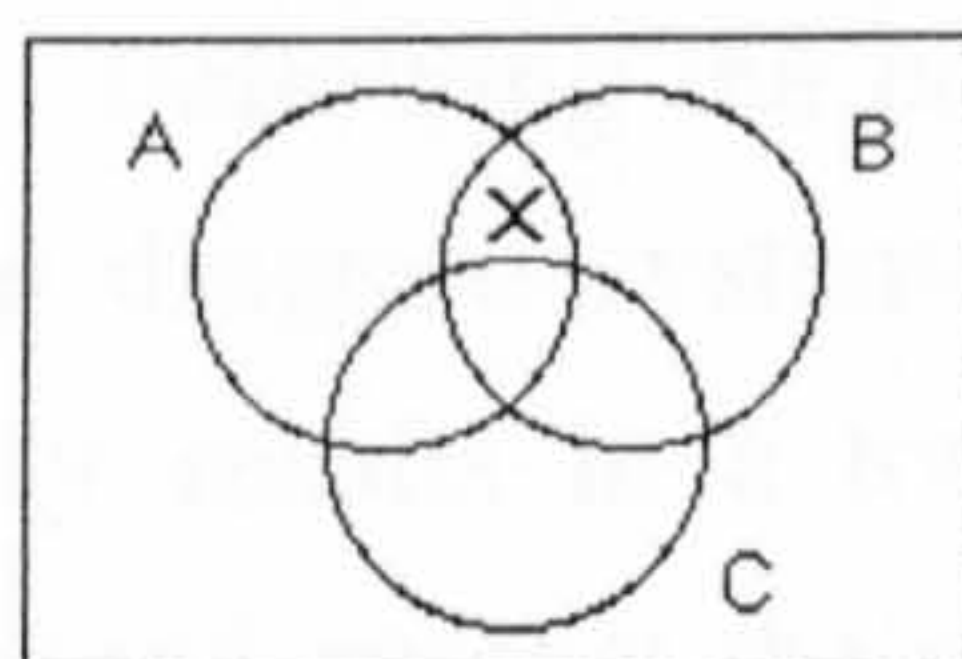


(b)

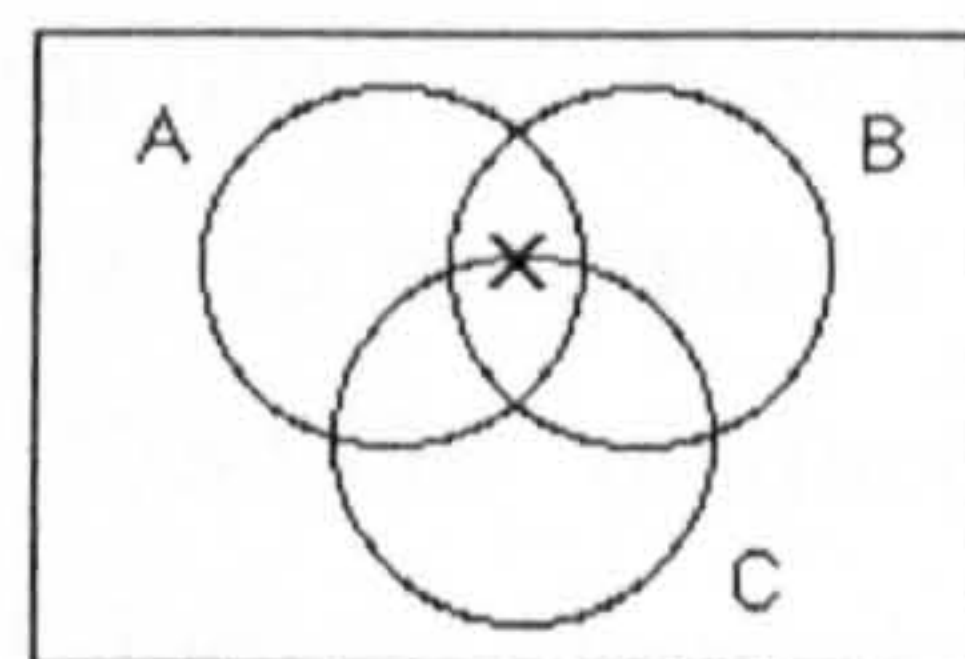


(c)

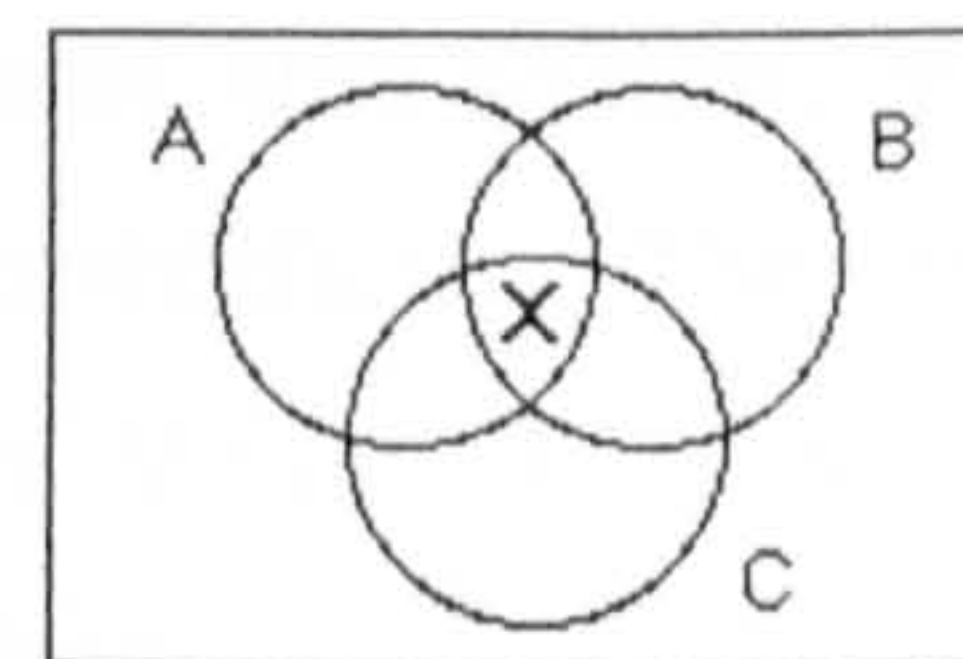
Assertions that there exist at least one case in a region (existential assertions) are represented by marking with an X a region which is not empty. The complementary nature of marking a cross for existence and shading out nonexistence mirrors directly the equivalencies: a) Some A are B  $\equiv$  Not (No A is B), and, b) Some A are not B  $\equiv$  Not (All A are B). Thus the following diagrams all indicate that Some A are B.



(d)

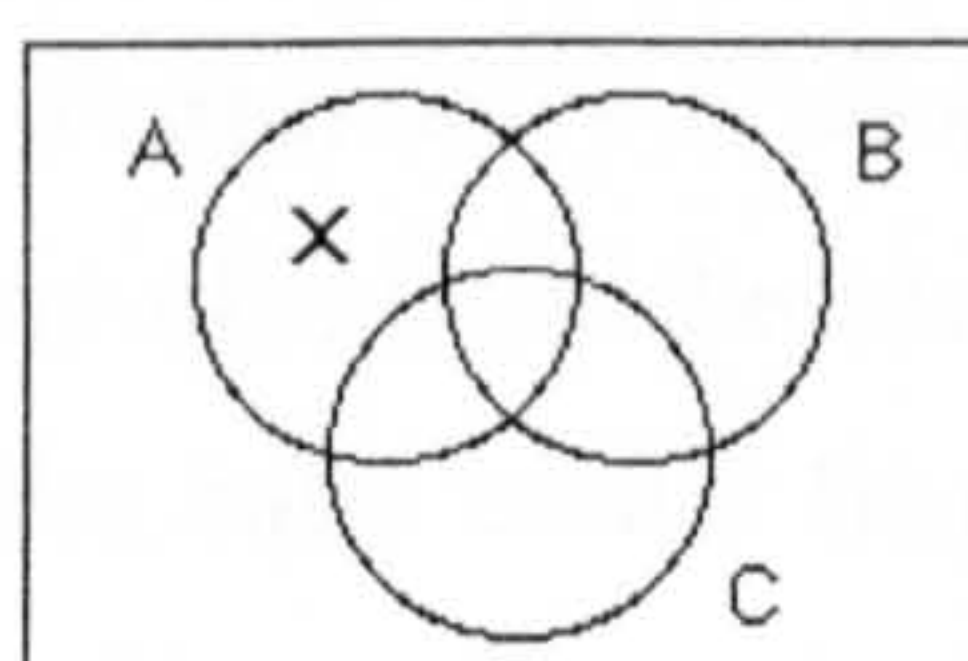


(e)

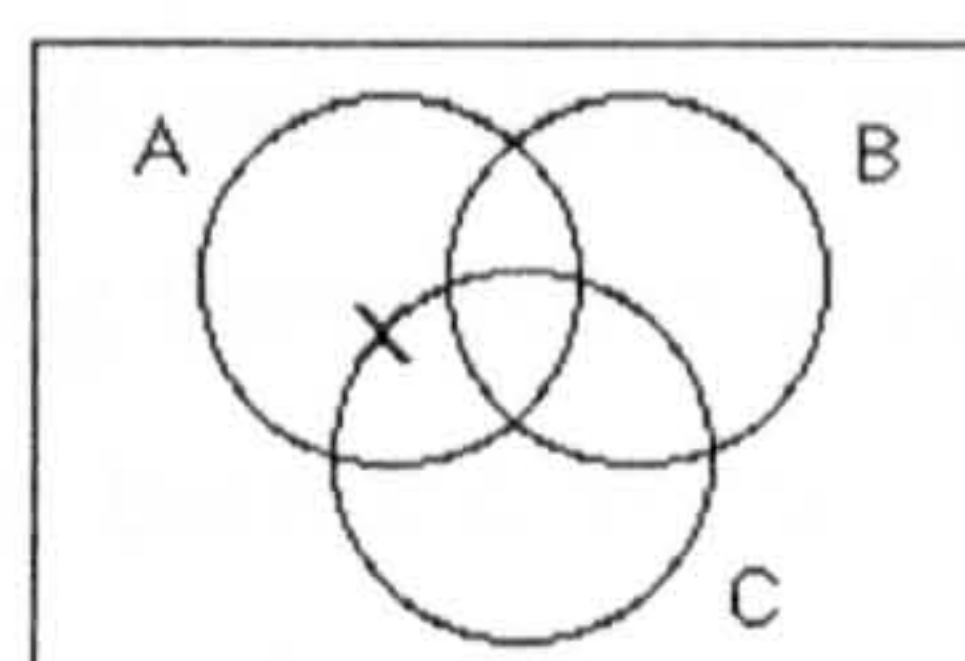


(f)

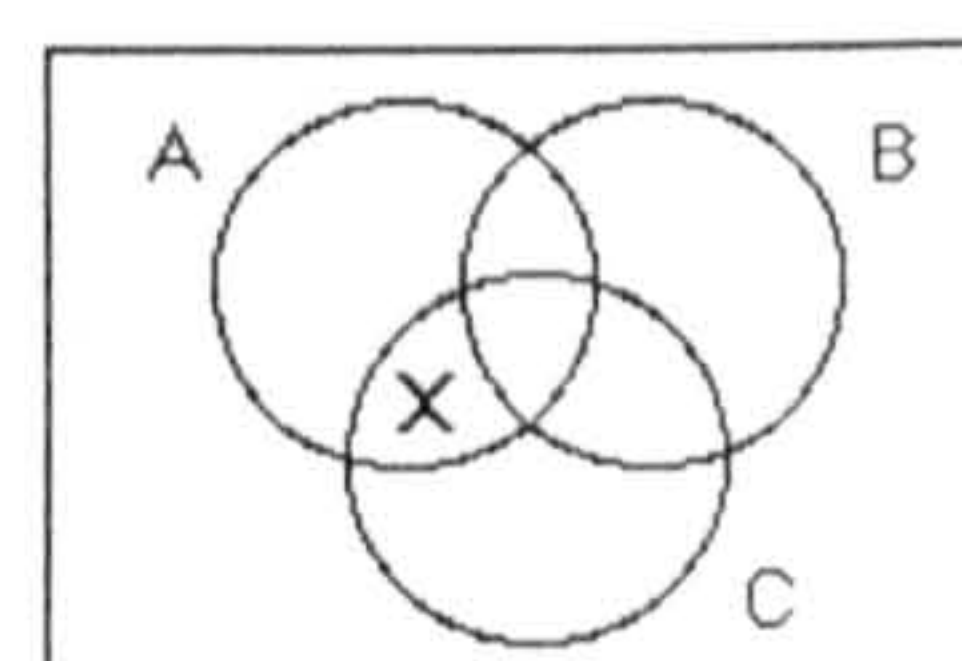
The difference here is in the additional information concerning the relationship between A and C. In (f) there is a clear assertion that Some A are C whilst in (d) there is a suggested implication that this is not so. In fact if the rules are adhered to then such an implication is not supported without appropriate shading so that it may be argued that (d) is quite acceptable. Nevertheless we adopt (e) as the conventional way of representing the statement Some A are B without further concern for C. In precisely the same way the following diagrams all signify that Some A are not B



(g)



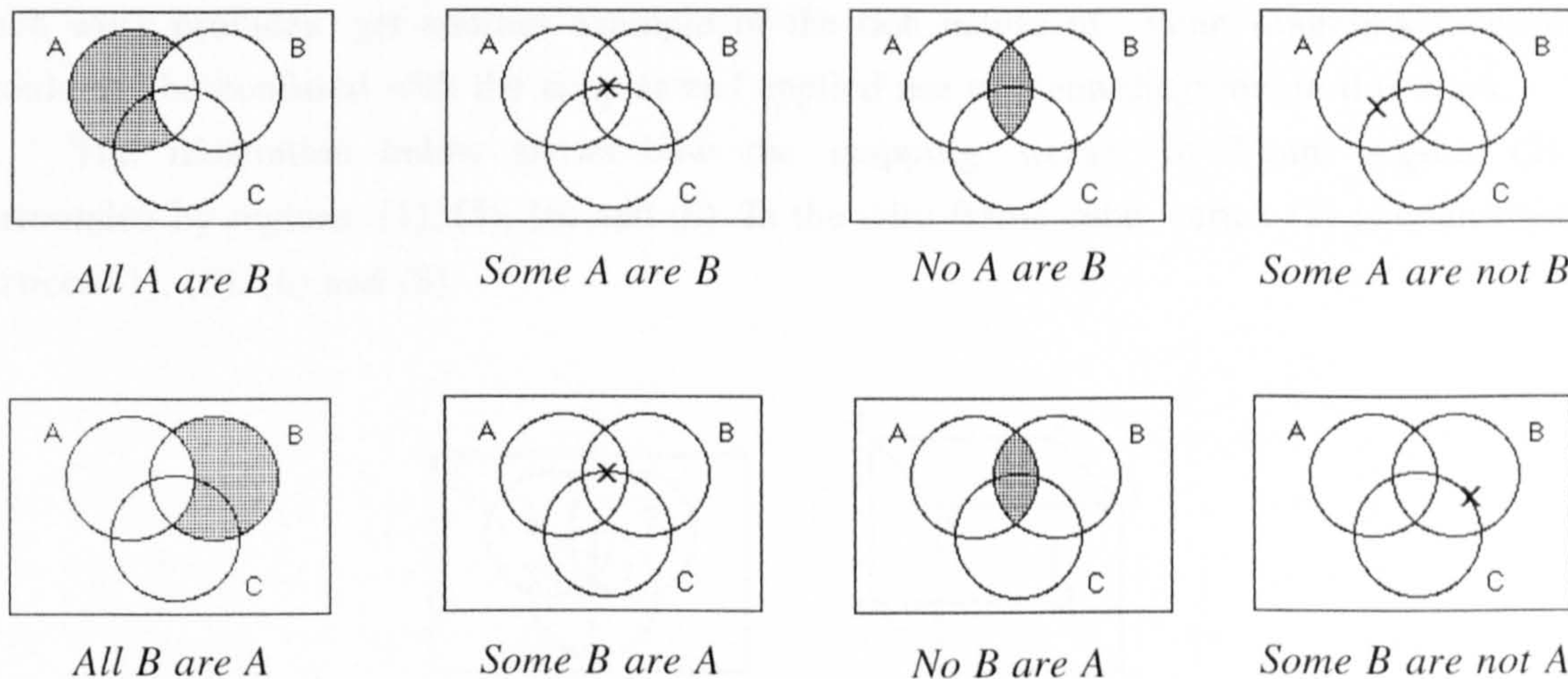
(h)



(i)

Here (g) also suggests (but again this is unsupported) that No A are C whilst (i) actually asserts that Some A is C. Diagram (h) is adopted as the conventional representation of Some A are not B. In total there are eight forms which the first premise of a syllogism can take.





Logically equivalent premises produce the same diagrams for example the diagram for Some Bs are As is the same as for Some As are Bs.

Following the counting argument used to show the office system, this section treats the Venn diagram systems in the same way to produce a measure of specificity. The counting finally results in a total number of 1784593 distinct diagrams in Venn. This number is measured against the same number of domain elements for each system.

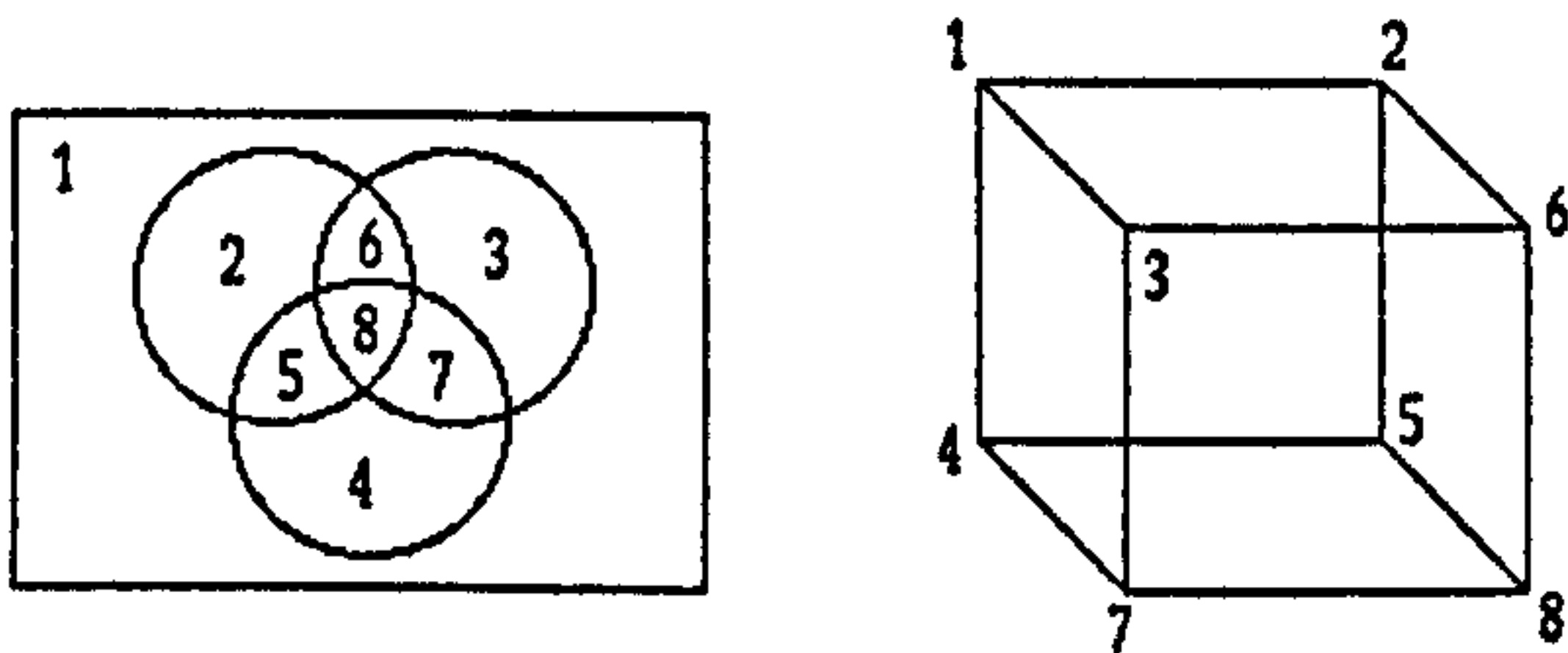
The count starts by assuming that an 'x' may be placed in the interior of any of the 8 regions or on any of the 12 arcs that separate those regions. The total number of possibilities is then  $2^{20}$ . Of course any of the 8 regions may be shaded giving a further  $2^8$  possibilities. In principle these could be combined as independent but earlier remarks concerning inconsistency, place a restriction on this and the counting is a little more complicated. Shaded regions prohibit 'x's from any of the immediately surrounding boundaries. This restriction makes the counting quite difficult to track. Different regions have different numbers of boundaries. Since those boundaries are also shared with other regions, these regions must also be tracked in the counting.

Fortunately a wire-frame cube has quite similar properties to the Venn diagram and for reasons discussed below, makes tracking these possibilities more straight forward. The parallel with the cube stems from the mapping between vertices of the cube with contiguous regions of the Venn diagram. The wire frame cube has appeared as a method for illustrating the syllogism (Stenning, 1993, December), but does not appear anywhere else used for counting states for other representations. Towards the end of this research it was interesting to find that mathematicians have developed the connection between Venn diagrams and their internal geometry way beyond the use made here. Generalized models of graphs associated with Venn provide elaborate illustration of combinatorics (Ruskey, 1997; Bultena and Ruskey, 1998).



Such work provides yet another example of the rich nature of Venn diagrams however it should not be confused with the simpler and applied use of Venn diagrams in this work.

The illustration below shows how the mapping works. In Venn, region (2) is surrounded by regions, (1), (5), (6) and (8). In the wire-frame cube, vertex (2) is connected to vertices (1), (5), (6) and (8).



Counting the number of states in the Venn diagram requires eight different scenarios. Each scenario takes a number of shaded regions and considers the patterns of available arcs and regions. These patterns are always symmetrical and are multiplied by the repetitions in symmetry.

The first two elements of the counting argument do not require the parallel with the wire-frame cube. If no regions are shaded, an 'x' can appear in any one of the eight available regions or on any of the 12 arcs. This yields  $2^{20}$  possibilities. The counting progresses by increasing the number of shaded regions assumed. There are eight ways to shade one region (there are eight regions). For each of these eight alternatives, 7 other regions and 9 arcs are left. These options add to  $2^{16}$  possibilities. These two figures account for the first and second entries in Table 6 below.

Table 6: The Venn counting argument.

Shaded regions	States
0.	$2^{20}$
1.	$8 \times 2^{16}$
2.	$5 \times 2^{15}$
3.	$19 \times 2^{11}$
4.	$237 \times 2^5$
5.	$19 \times 2^6$
6.	$5 \times 2^5$
7.	16
8.	1
Total	1784593

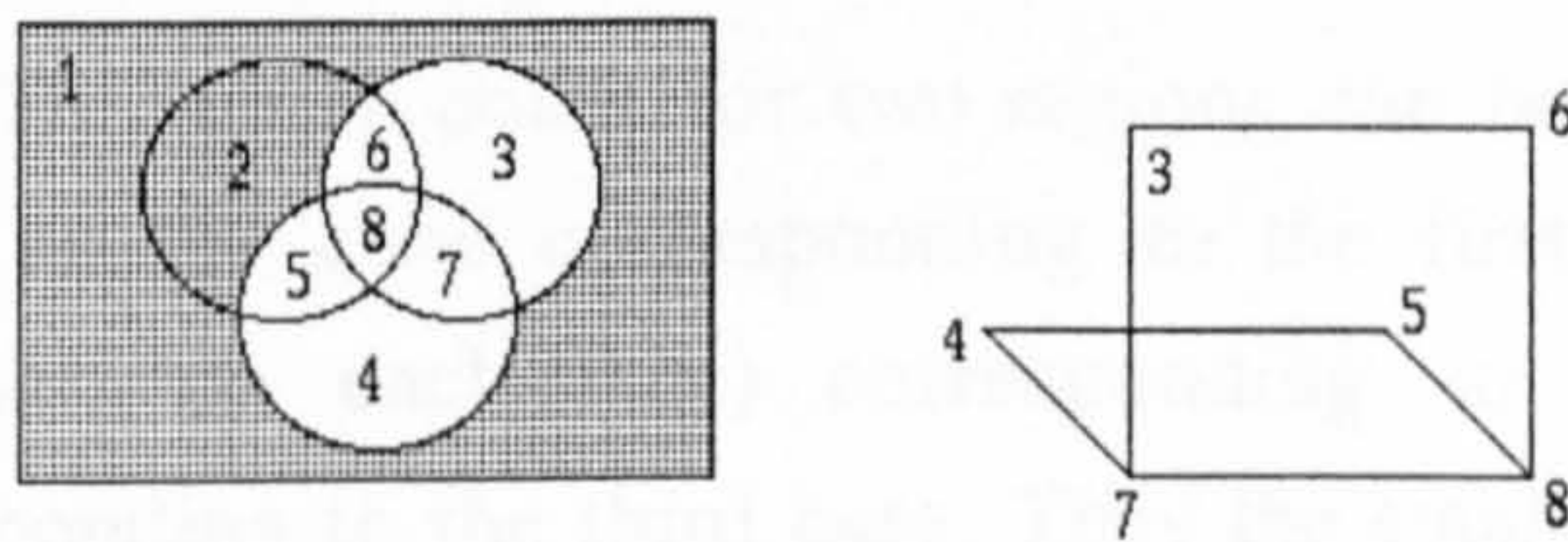
The cube illustrates the allowable patterns of 'x' annotation for patterns of region shading. The diagram could be shaded with two regions in three separate patterns. Each pattern is repeatable. The first kind of two shading pattern is illustrated by the edges on a



cube. Edges are the imaginary straight lines that link the vertices of a cube. Every cube has twelve of these edges. Any argument that works for one edge must apply to all others. The total number of patterns possible from this 'edge' configuration is the multiple of those possible for one edge and the number of edges in the cube.

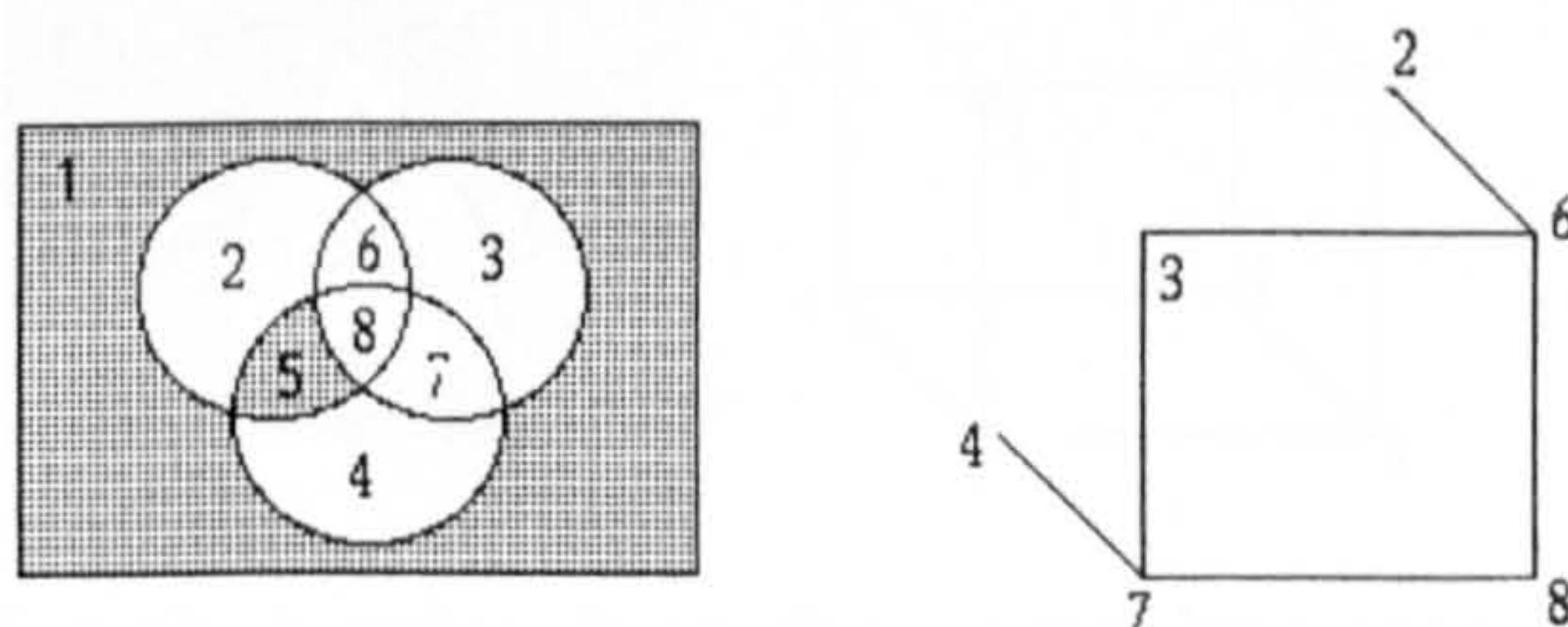
### Two Shaded Regions

The first prototypical two shaded region configuration is shown with the cube and the Venn diagram below. The diagram shows regions 1 and 2 shaded. The equivalent wire-frame cube representation removes vertices 1 and 2 from the illustration. Removing these vertices also removes the associated edges. These are potential locations in the Venn diagram for an 'x' to be placed. They are removed because they represent edges that are prohibited from 'x' annotation in this case. The prohibition stems from shading in the immediately adjacent regions.



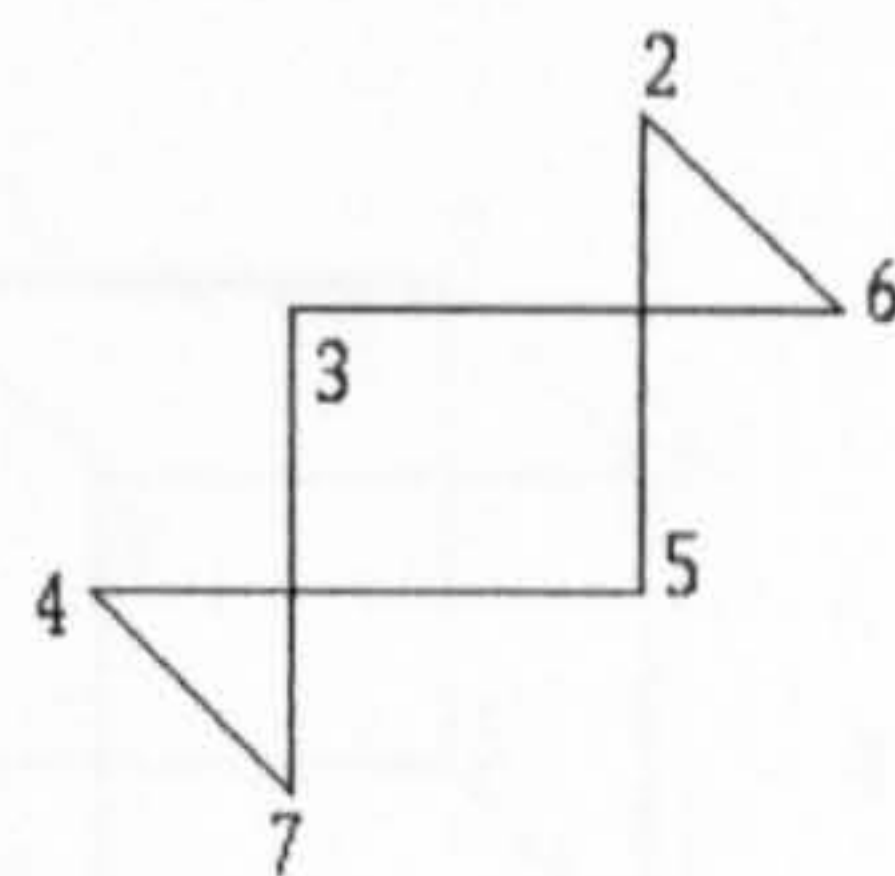
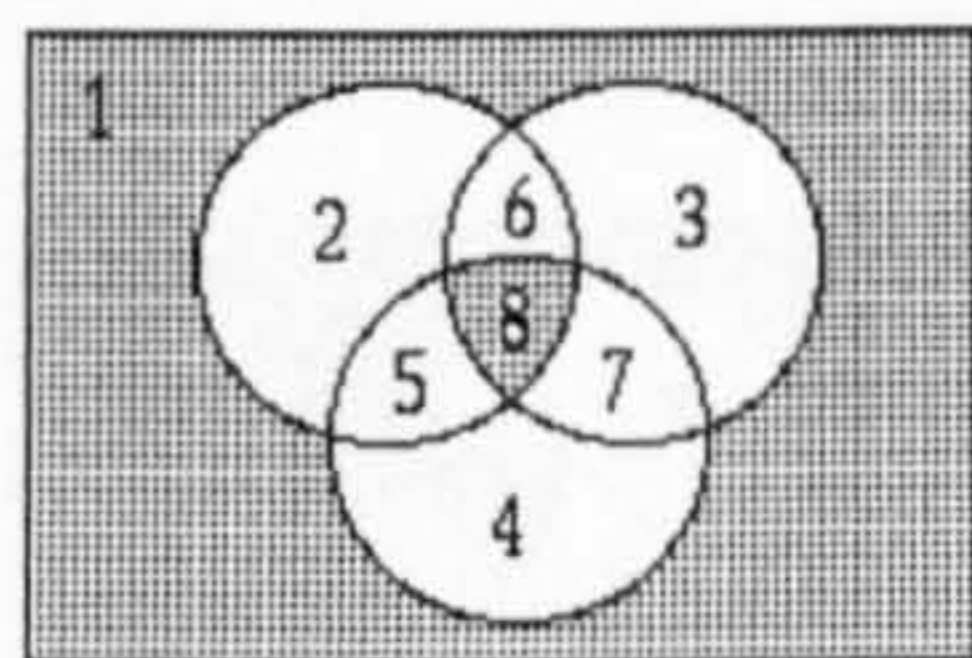
When regions one and two are removed, six vertices and seven edges of the cube are left. These represent six regions and seven arcs in the Venn diagram. This will yield  $2^{13}$  different figures resulting from the combination of multiple "x"s. The cube picture illustrates that there are 12 of these edges and hence 12 multiples of the possibilities shown for one edge. The total number of states for the 'edge type' two shaded region configuration is therefore  $12 * 2^{13}$  possibilities.

In the second case we make use of the fact that there are three regions of distance two from region 1 namely regions 5, 6 and 7. The corresponding picture for region 5 and region 1 is shown.



In this situation six vertices and six edges remain yielding  $2^{12}$  possibilities for an 'x'. In the third case the only remaining choice is the diagonal region 8 or in other words the region which when seen in relation to region 1 maps onto the diagonal of the cube. This yields the single picture:





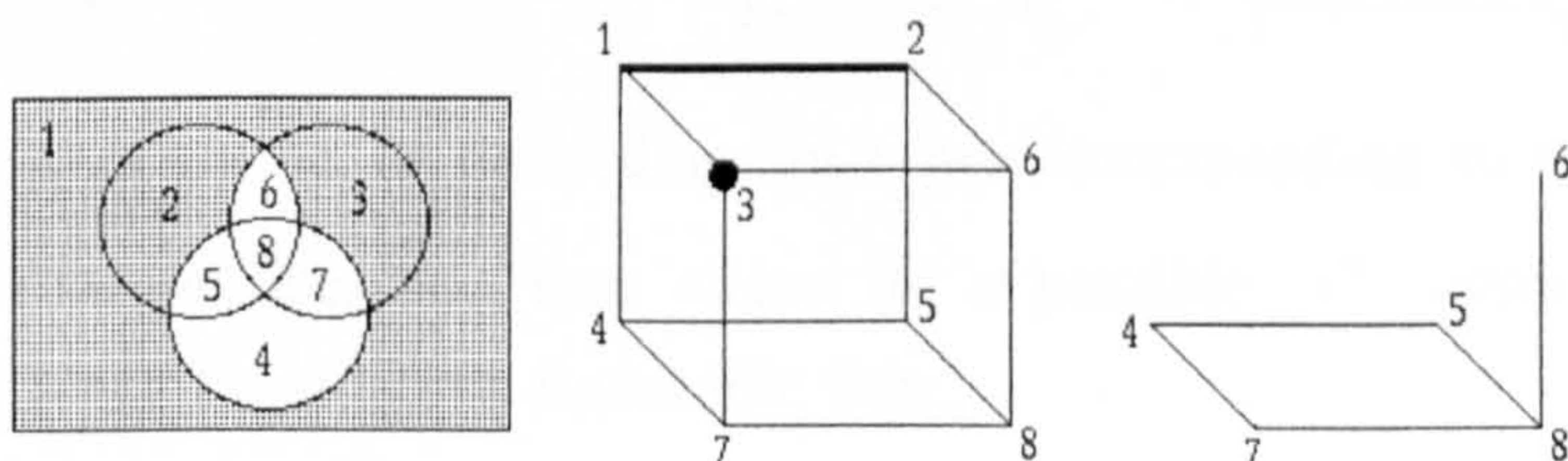
Here there are again six vertices and six edges and therefore  $2^{12}$  possible 'x' markings. To add all these possibilities requires some care. We chose vertex 1 to start with and there are eight such choices but some duplication will occur. The number of distinct ways of shading two regions is given by  ${}^8C_2 = 28$  calculated from the following, where 'x' is the number of options and 'y' the number of alternatives.

$${}_x C_y = \frac{X!}{Y! * (X - Y)!}$$

The whole count for two regions can be broken down on the basis that there are 12 edges on the cube corresponding to the first case above then 12 diagonals of faces (two diagonals on each face) corresponding to the second case and four cube diagonals corresponding to the third case. Thus the total number of 'x' markings when two regions are shaded is  $12*2^{13} + 12*2^{12} + 4*2^{12} = 5*2^{15}$ .

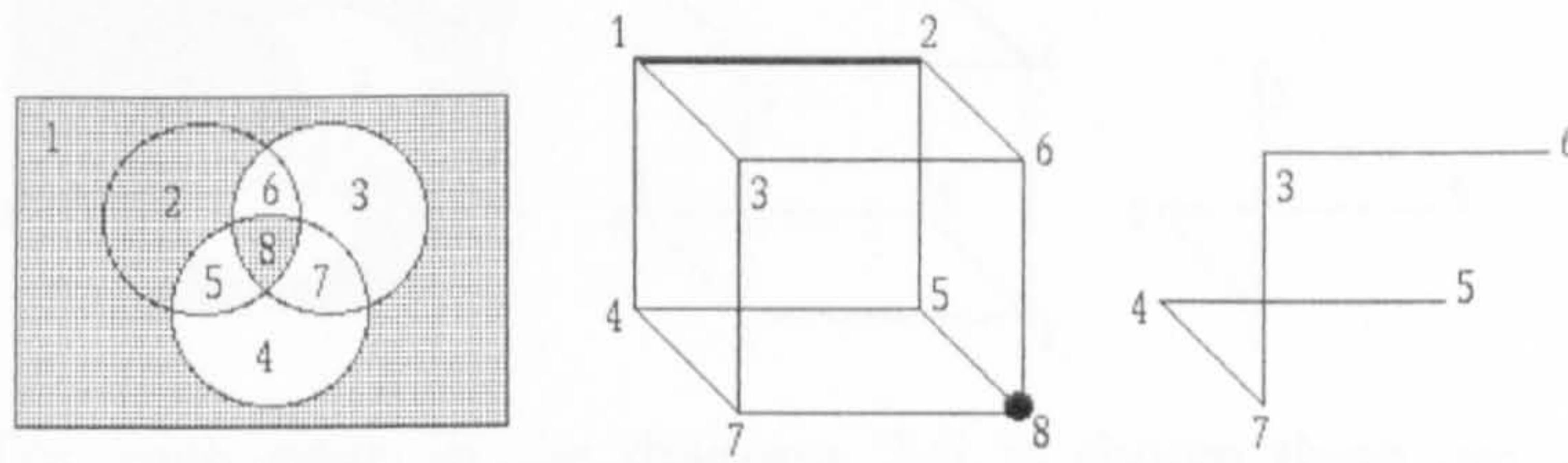
### Three shaded regions

Of the  ${}^8C_3 = 56$  ways of shading three regions there are three generically different types. Again these are best illustrated by reference to the cube picture. In the first case the cube subtypes look like the following:

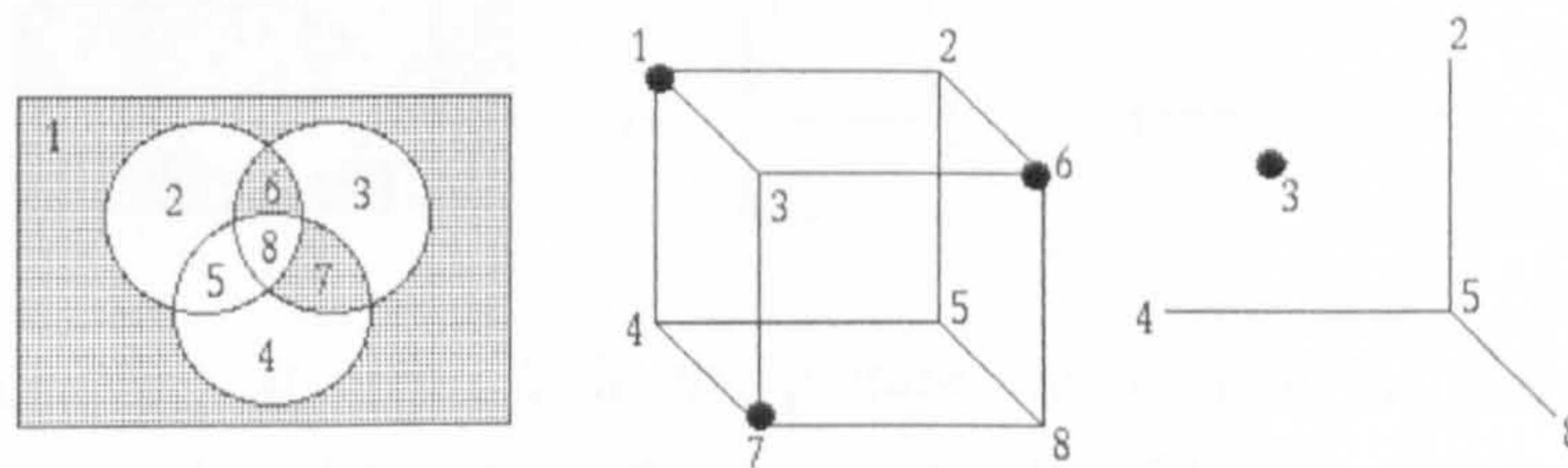


For each edge it is the same however edge 1—2 is shown, there are four nearest neighbours. Adding all 12 edges gives a total of 48 configurations but each is counted twice. Thus there are 24 such patterns and each leaves five vertices and five edges as possible locations for an 'x' totaling  $24*2^{10}$  opportunities for 'x' annotation. In the second case for each edge, again 1—2 is shown, there are two diagonal opposites.





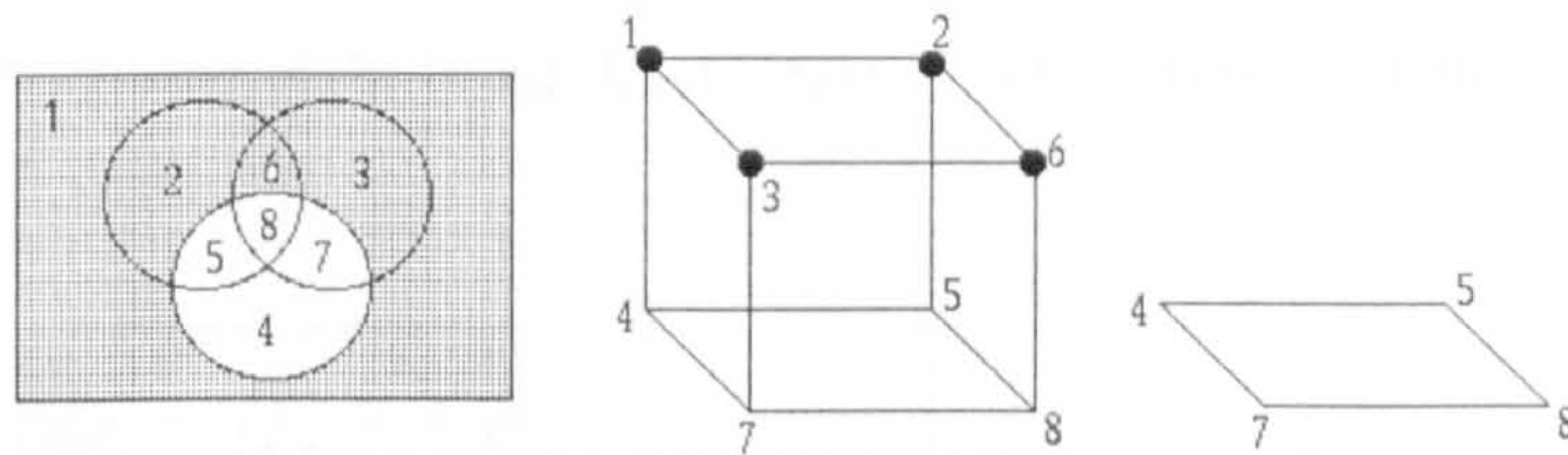
Again there are 24 of these patterns though in this case none are duplicated and each leaves five vertices and four edges as possible sitings for an 'x' thus yielding  $24 \cdot 2^9$  possibilities. In the third case there are eight ways to choose three non-contiguous vertices.



There are exactly eight ways of choosing three noncontiguous vertices and each leaves five vertices and three edges adding to  $8 \cdot 2^8$  possibilities. In total there are  $24 \cdot 2^{10} + 24 \cdot 2^9 + 8 \cdot 2^8 = 19 \cdot 2^{11}$  possible 'x' markings.

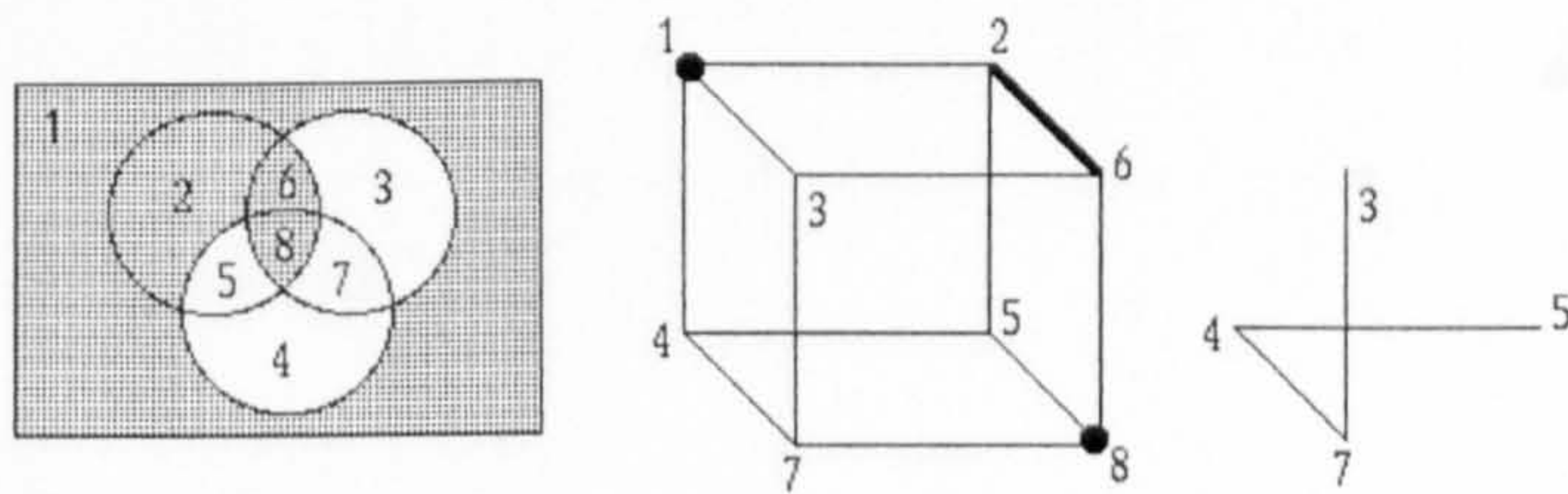
#### Four shaded regions

For four shadings there are  ${}^8C_4 = 70$  patterns which occur in six groups.

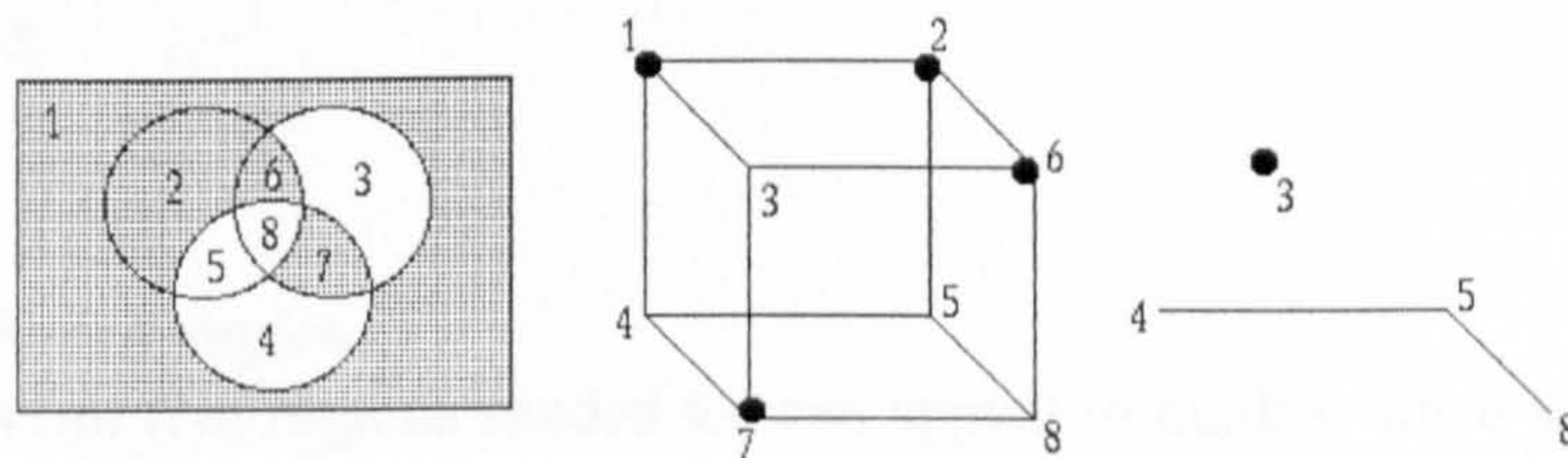


There are six of these diagram types corresponding to the six faces of the cube. Each leaves four vertices and four edges for a possible 'x' yielding  $6 \cdot 2^8$  possibilities. In the second place the diagram looks like this:

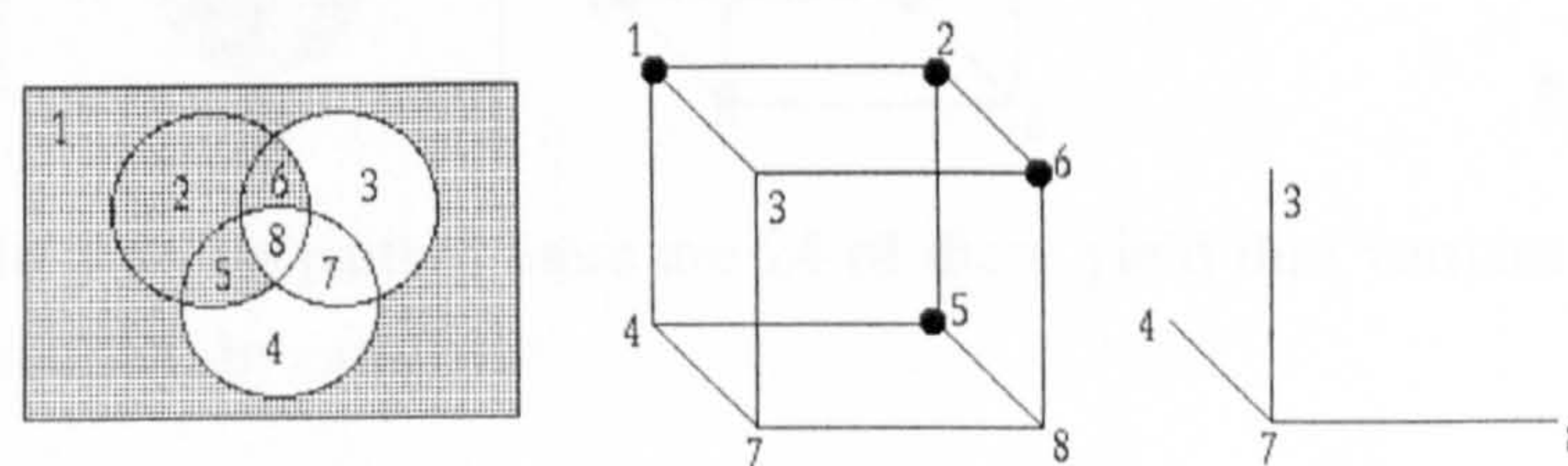




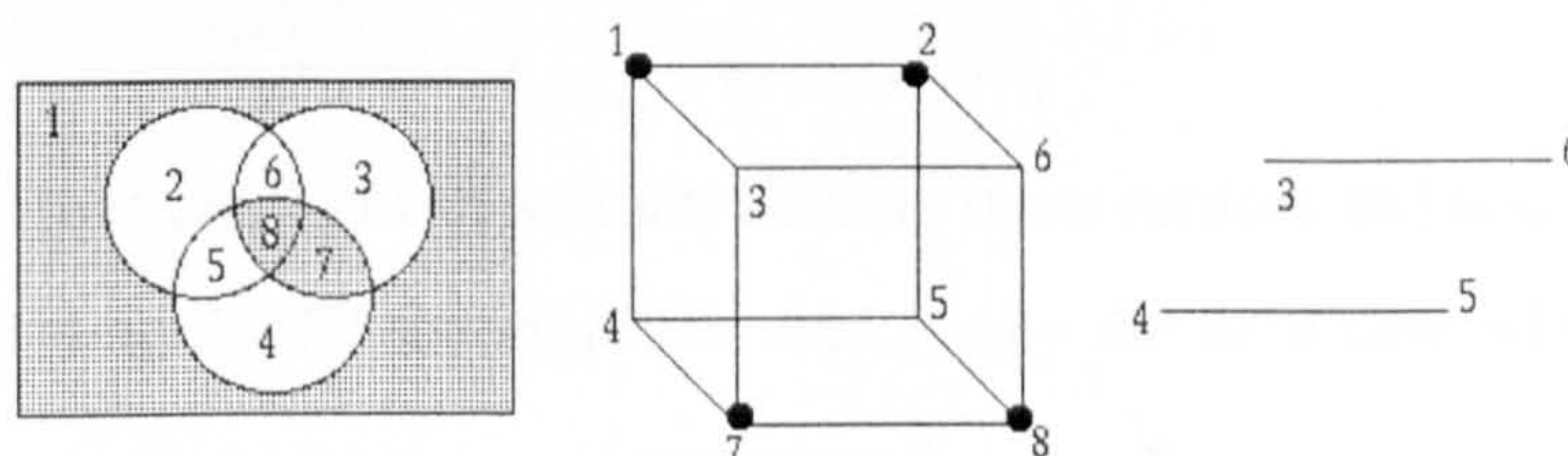
For each edge in the diagram 2-6 is shown there are two ways of completing this pattern giving a total of 24 such configurations. Each leaves four vertices and three edges yielding  $24 \cdot 2^7$  possibilities. In the third case the diagram looks like this:



Looking at vertex 2 in the picture we see that there are three such patterns arising. Since there are eight vertices there are a total of 24 such patterns. Four vertices and two edges remain for a possible 'x' and so  $24 \cdot 2^6$  opportunities for annotation exist.

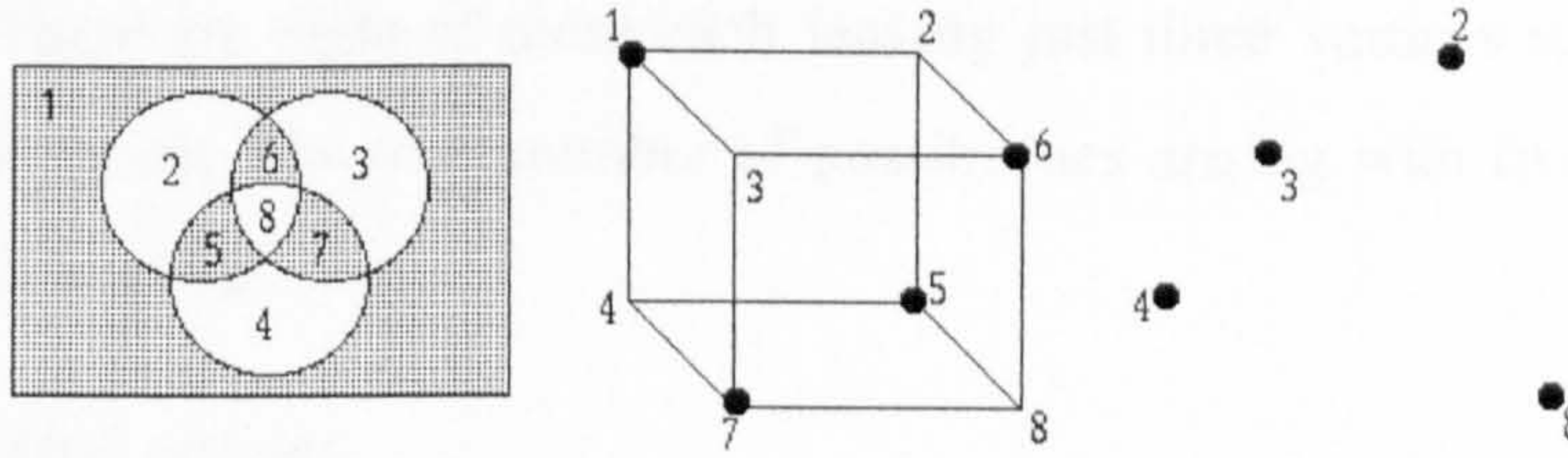


There is one pattern like the one shown here around each vertex (vertex 2 in the picture), giving eight in all. Each leaves four vertices and three edges so  $8 \cdot 2^7$  possibilities. In the fifth case:



Edges occur in opposite parallel pairs and there are six such. Each configuration of this type leaves four vertices and two edges yielding  $6 \cdot 2^6$  possibilities. In the sixth place:

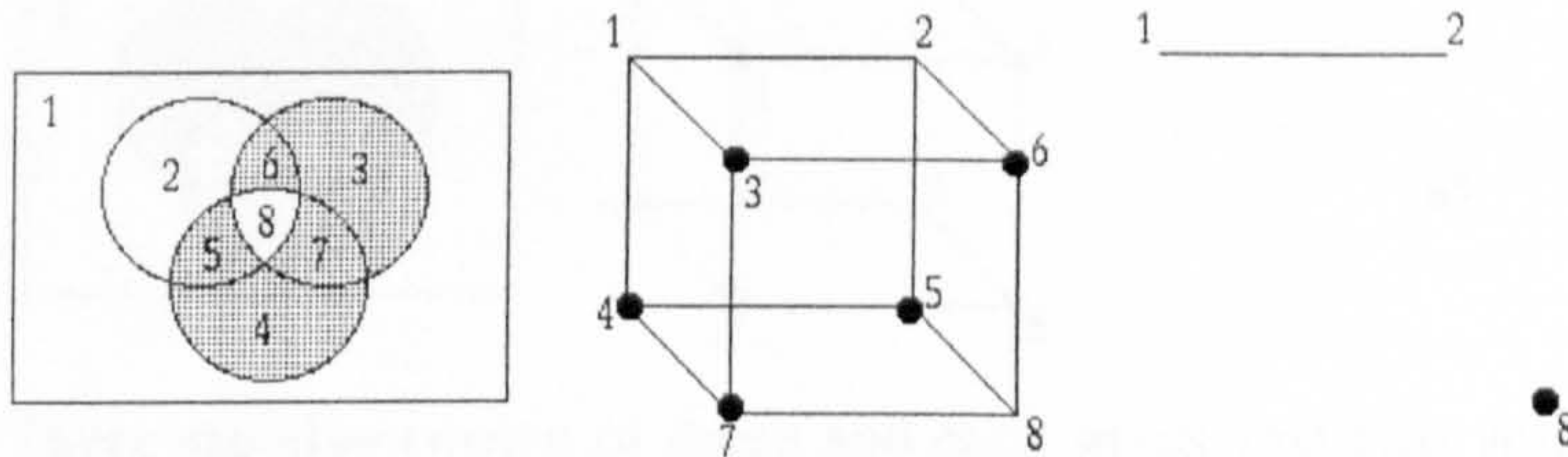




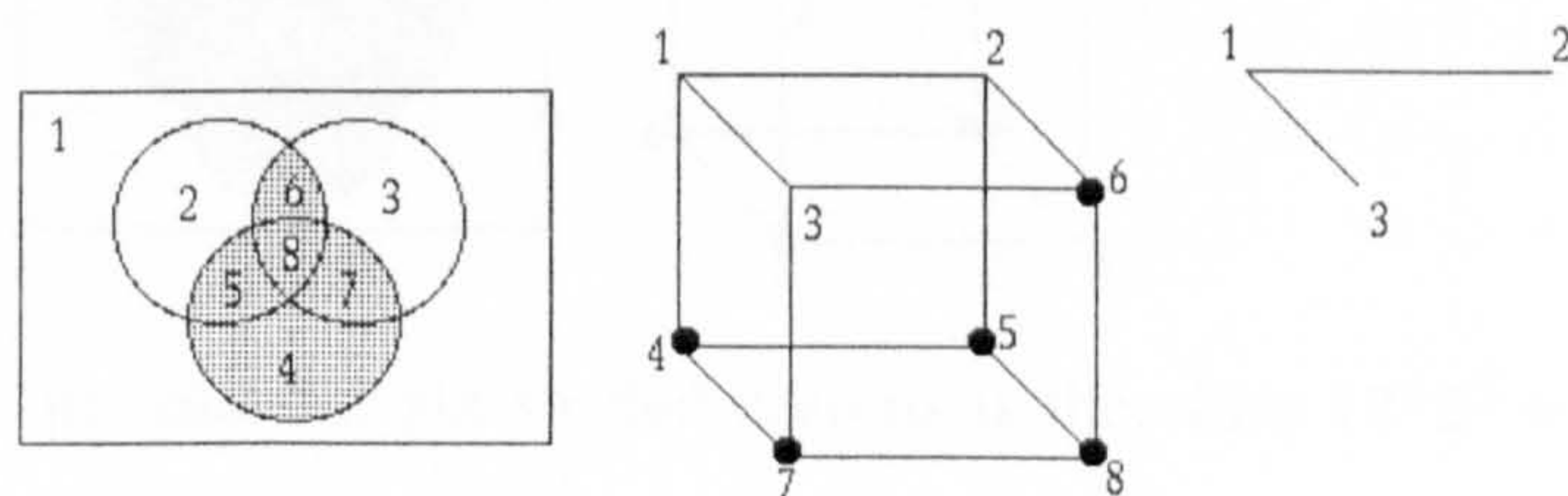
The final possibility consists of diagonals of two opposite faces giving a further two configurations. Each leaves four vertices and no edges so  $2 \cdot 2^4$  options. The combined total of possible 'x' sites arising for four 'x's are  $6 \cdot 2^8 + 24 \cdot 2^7 + 24 \cdot 2^6 + 8 \cdot 2^7 + 6 \cdot 2^6 + 2 \cdot 2^4 = 237 \cdot 2^5$ .

#### Five shaded regions

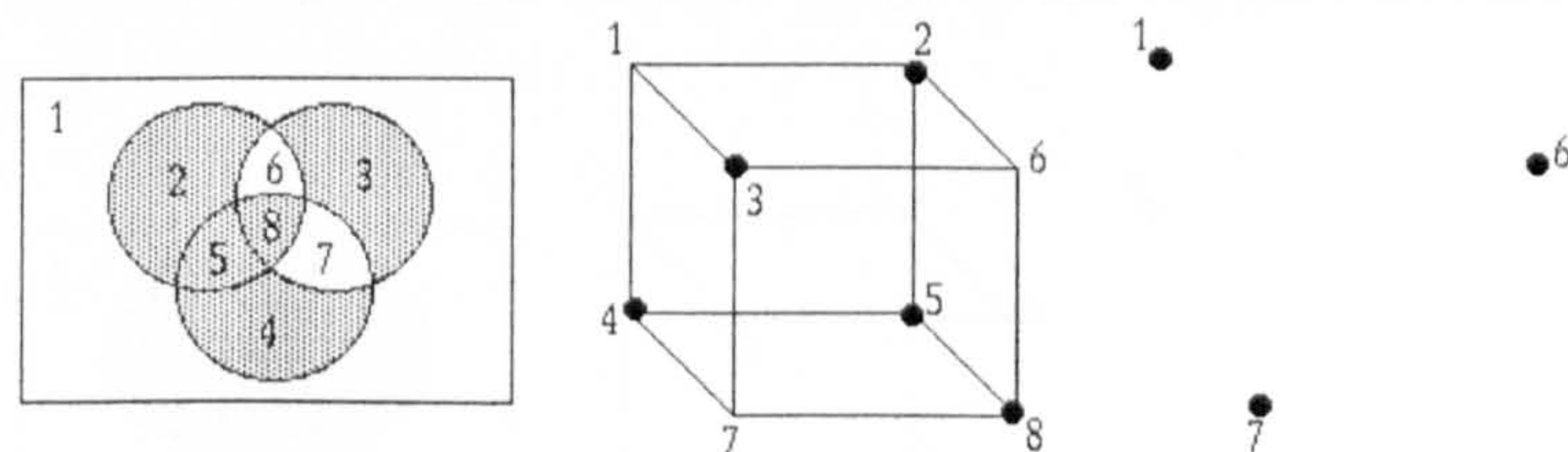
With five regions shaded we can appeal to duality since shading  $n$  regions is equivalent to not shading  $8-n$  regions! The three generic patterns are as follows:



In the first pattern there are 24 of these yield three vertices and one edge yielding  $24 \cdot 2^4$  opportunities. In case two:



There are 24 of these each offering three vertices and two edges as possible sites for an 'x', yielding  $24 \cdot 2^5$  opportunities. And finally the third case with five regions:

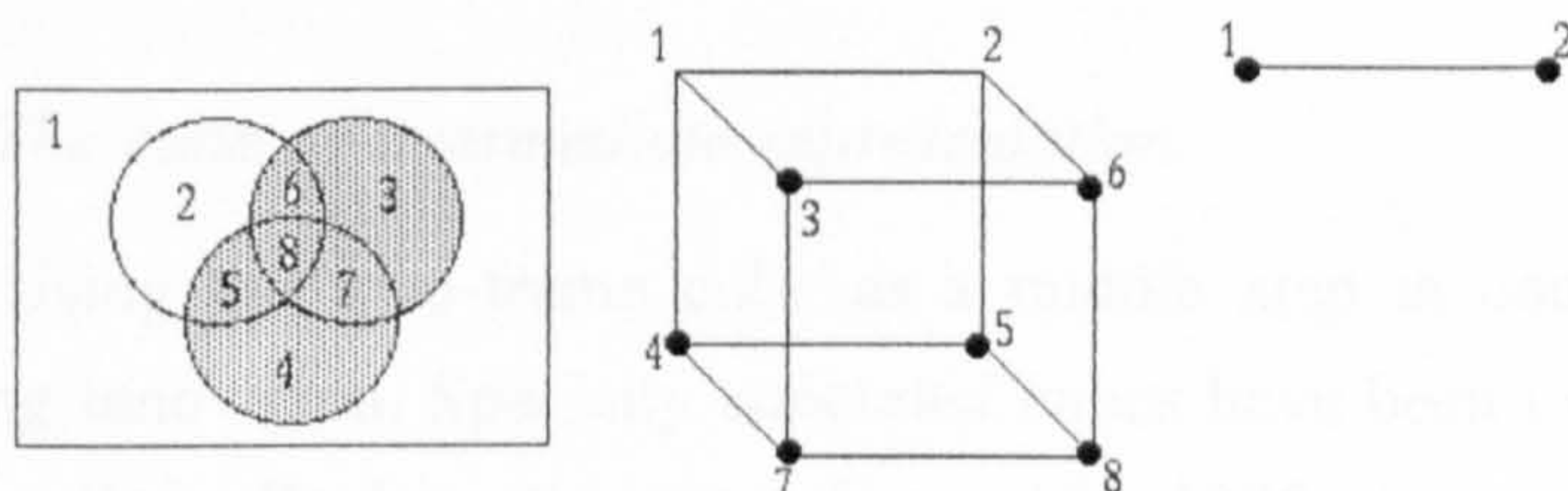




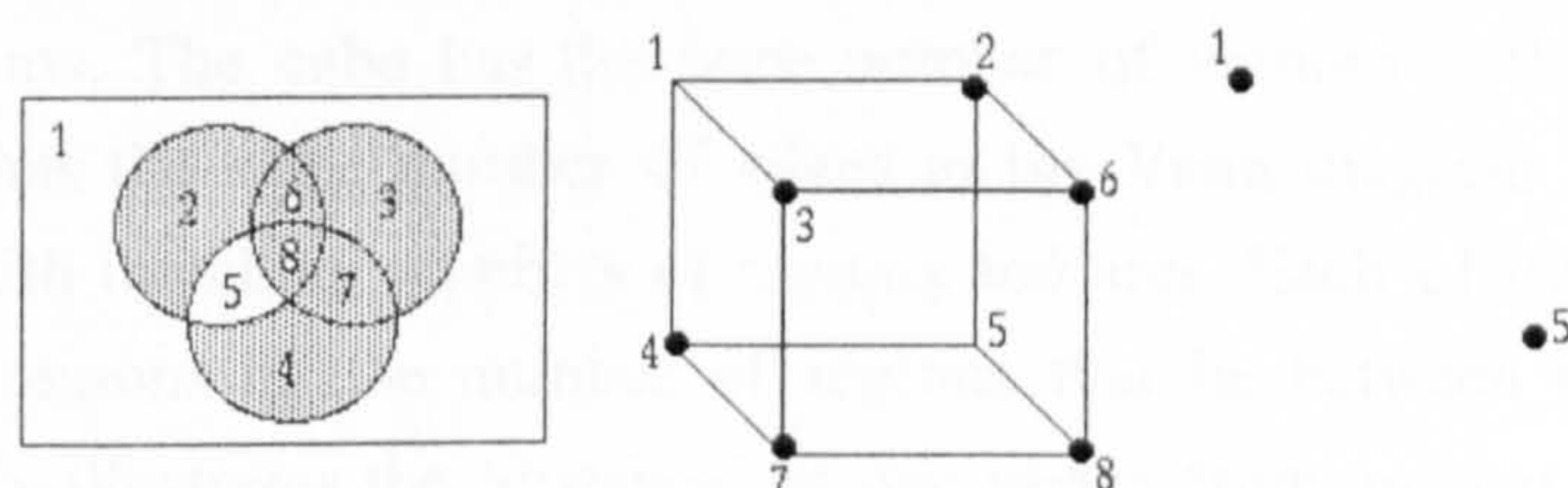
There are eight of these each leaving just three vertices so allowing  $8 \cdot 2^3$  opportunities for annotation. The total number of possibilities arising with five 'x's then is  $24 \cdot 2^5 + 24 \cdot 2^4 + 8 \cdot 2^3 = 19 \cdot 2^6$ .

### Six shaded regions

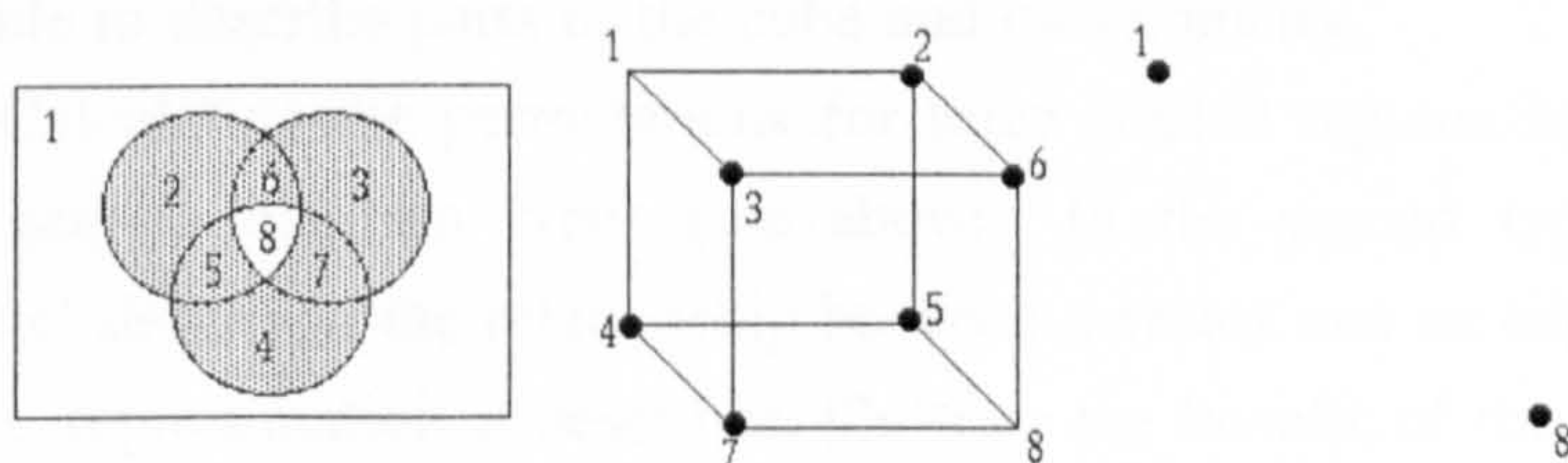
With six regions shaded there are again three patterns:



Each of these twelve yields two vertices and one edge.



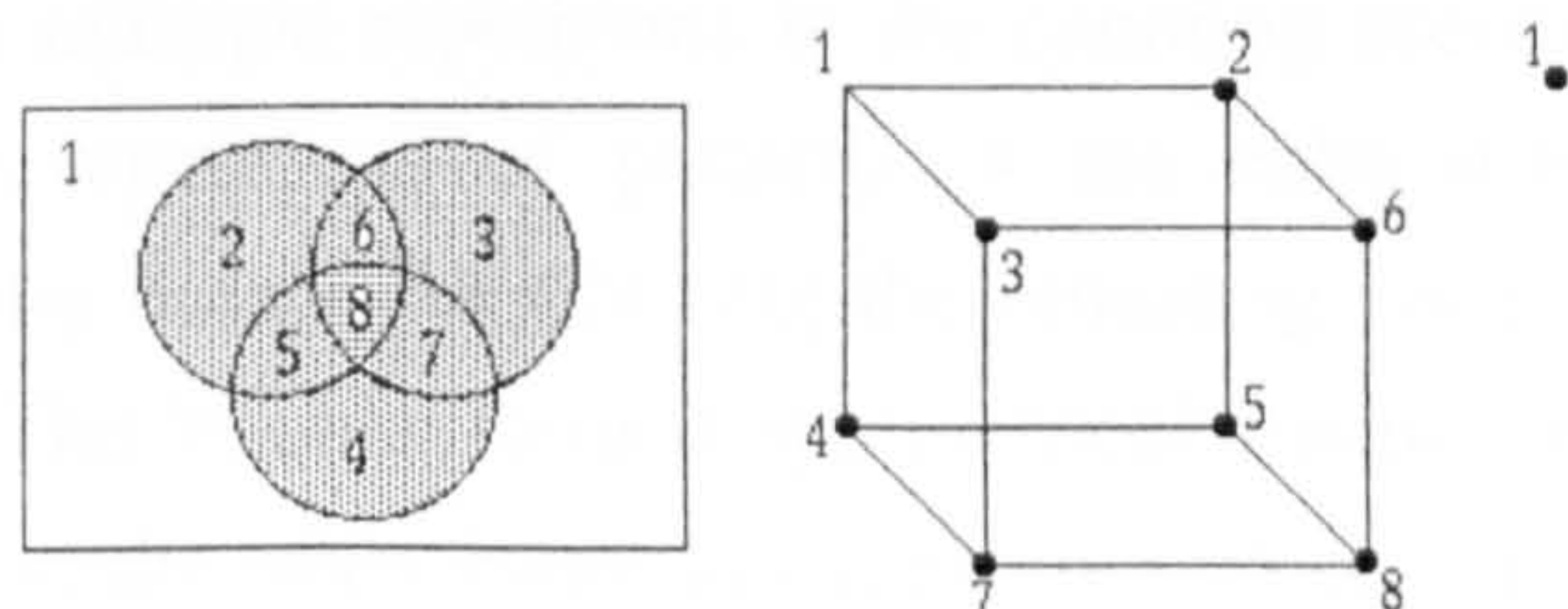
There are also twelve of these and each gives just two vertices. The remaining four also give just two vertices.



The total for six shaded regions is therefore  $12 \cdot 2^3 + 12 \cdot 2^2 + 4 \cdot 2^2 = 5 \cdot 2^5$ .

### Seven and eight shaded regions

There are just eight ways to shade seven regions this diagram shows only region one unshaded.



Each example of shading seven regions leaves only a single vertex. A further 16 possible 'x' markings arise this way. Finally if all eight regions are shaded then there can be no 'x's. These values are summarized in Table 6, above and the total number of states reaches 1784593. Most of these are of no use in the description of the 64 syllogisms. The Venn representation system is far more expressive than is needed. It can in fact express any first order relationship between the three sets of which there are 1784593.

#### *4.3.1 The cube as intermediate representation.*

Using the wire-frame cube as a middle step in counting the states in Venn was an exciting innovation. Specially annotated cubes have been used as an independent method for proving the syllogism (Stenning, December, 1995) but have not been used as intermediate representations as they are used here.

The wire-frame cube is useful because it is similar in form to the Venn and Carroll diagrams. The cube has the same number of vertices as the Venn diagram has regions. The cube has the same number of edges as the Venn diagram has arcs. The similarity does not end with matching numbers of regions and arcs. Each of the regions in Venn corresponds to other regions by the number of regions that lie between them. In the same way, the cube directly illustrates the 'distance' of one vertex from another. The Venn diagram is a modified wire-frame cube diagram in which lines are mutated but not deleted or added.

The benefit of the cube as an intermediate representation depends on the language available to describe parts of the cube and its symmetry.

Calculating the permutations for three shaded regions in the Venn diagram generated three generic diagram types (see above). In the second type, the expression, 'diagonal opposite' described the relationship between a vertex and an edge in the cube. This language and the representation it describes, illustrate the benefit of the cube. There is no equivalent language to describe the corresponding pattern of shading that refers to the Venn diagram.

Illustrating the equivalent shading pattern in a Venn diagram requires the regions are numbered in a relatively arbitrary way.

The wire-frame cube embodies a very convenient language to support the process of counting possible permutations. The internal constraints of the Venn system acquire more obvious and familiar names when represented as a cube. Symmetries in the cube patterns enable multiple repetitions in the counting process. The repetitions in counting correspond to visible repetitions of patterns in the cube diagram. The cube provides a more obvious checking device for tracking the repeating patterns.

The Venn system is symmetrical around the three central axes of the three circles. The cube, on the other hand, has four lines of symmetry per face, six per vertex, two per edge and more through cube diagonals and rotations around vertices. The cube has at least twelve



symmetries whereas the Venn diagram has only three obvious lines of symmetry. Symmetry makes the computation of multiple patterns more straight forward by implying repetition in the counting. Hammer (1995) develops a theoretical explanation for this use of symmetry indicating that its main function is to reduce the number of cases needed for effective calculation.

4.3.2 Reducing the expressivity of Venn.

A system can be honed down by adding more constraints, until the system is only capable of a very limited degree of expression. This process has no definite end except when the states possible are equal to the material communicated. This lack of any *natural* end to the limiting process suggests the extra rules do not alter the benefit of the systems for the learner.

Each of the three circle systems are treated to one extra rule that limits the expression possible. This rule is that only the tokens required to illustrate two premises should be counted. This still allows for many permutations that do not represent states needed for the syllogism. However the rule does reduce the total number of permutations. Since there are only ever two premises and one conclusion, then in Venn, numbers of star and shading annotations will never be more than three (one for each term). With this constraint there are  $28 + 8*16 + 190 = 346$  configurations (as shown in Table 7 below).

Table 7: Expression in Venn with just two premises.

Annotation	Count	States
2 x.	$^{20}C_2$	190
2 shaded regions	$^8C_2$	28
1 shaded region & 1 x	$8*16$	128
Total expressivity.		346

Note: The  $^xC_z$  notation refers to equation for summing states with x choices between z alternatives.

There are various other ways of honing the graphical system down even further so that the conventions allowable are virtually no more flexible or expressive than the states of the syllogism itself. One such depends on the idea that no statement in the syllogism includes all three terms. On this argument, annotating the central region while illustrating the premises is not possible. The implication derived from this extra rule is valid. Nevertheless, in some situations combining certain kinds of premises, reading the conclusion from the diagram does require annotations to occupy the central region.

This restriction to the system and others like it show that by adding extra interpretation restrictions, systems can be reduced in their capacity to express. Assessing such systems with the information enforcement metric would clearly give them more promising scores than



their relations without the extra rules. A literal interpretation of the metric's predictions would lead to the expectation that the most constrained system would be best for learners. However, as these new rules are added, the representation also becomes more complex. It is more natural to count expressivity using the whole system, than to count a system with an arbitrary set of additional rules.

#### 4.4 The Edinburgh-Euler system.

The Euler circle system was explained in Chapter 2. In this section we are primarily interested in applying the specificity or information enforcement metric to the Edinburgh Euler system. Diagrams that are representative of each of the possible premise types are given in the Figure 12.

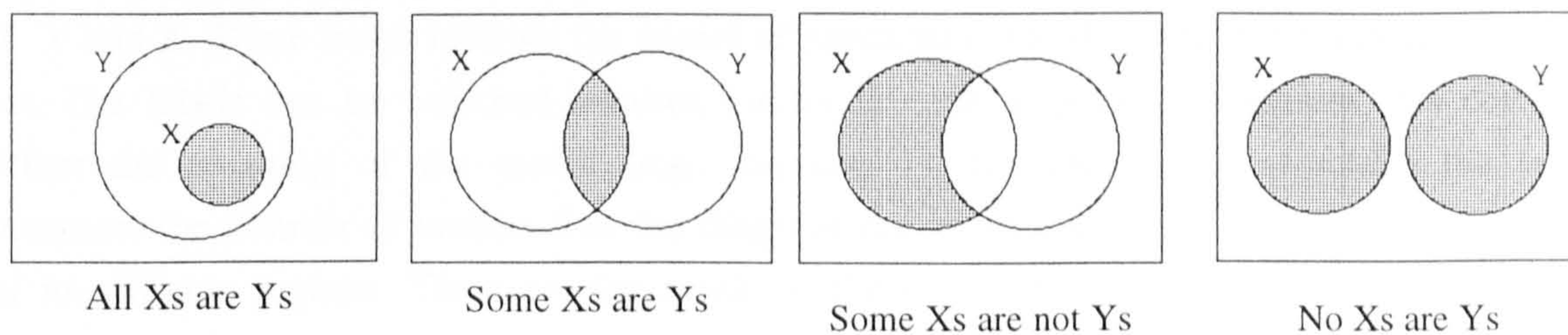
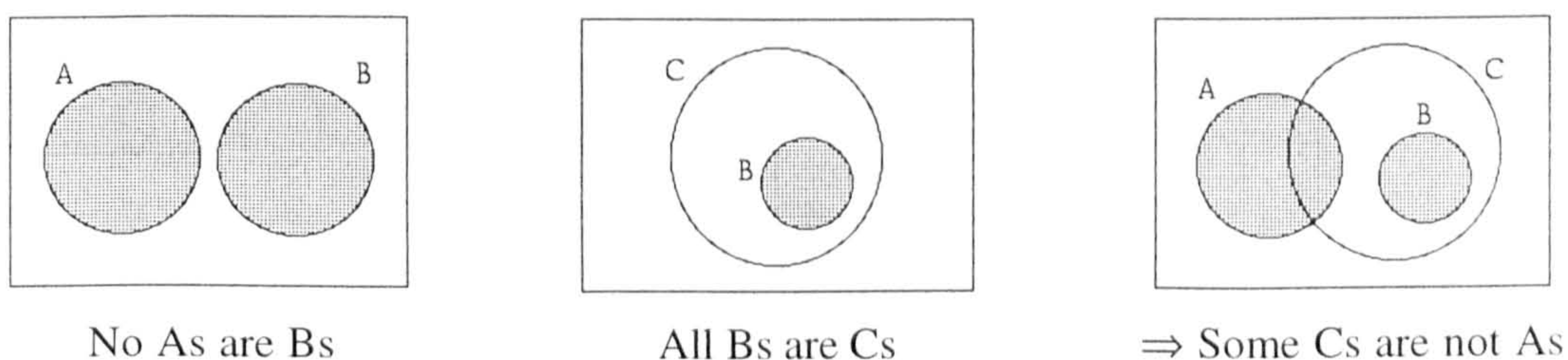


Figure 12: Premises in Edinburgh Euler.

In the first diagram the shaded X circle is contained within the Y circle indicating All Xs are also Ys. In the second diagram shading represents individuals that exist (in this case some of the Xs are Ys). In the third diagram the shaded area represents individuals which are Xs but not Ys. The intersection is not shaded since we do not know whether any Xs are Ys. In diagram four there are individuals in both X and Y but the X and Y regions do not intersect. The combination or registration of premise types is given by overlaying B term circles and by the development of as many regions as are consistent with the original premises. If there are any regions that remain nonintersected from this process, they become the subject of a valid conclusion.



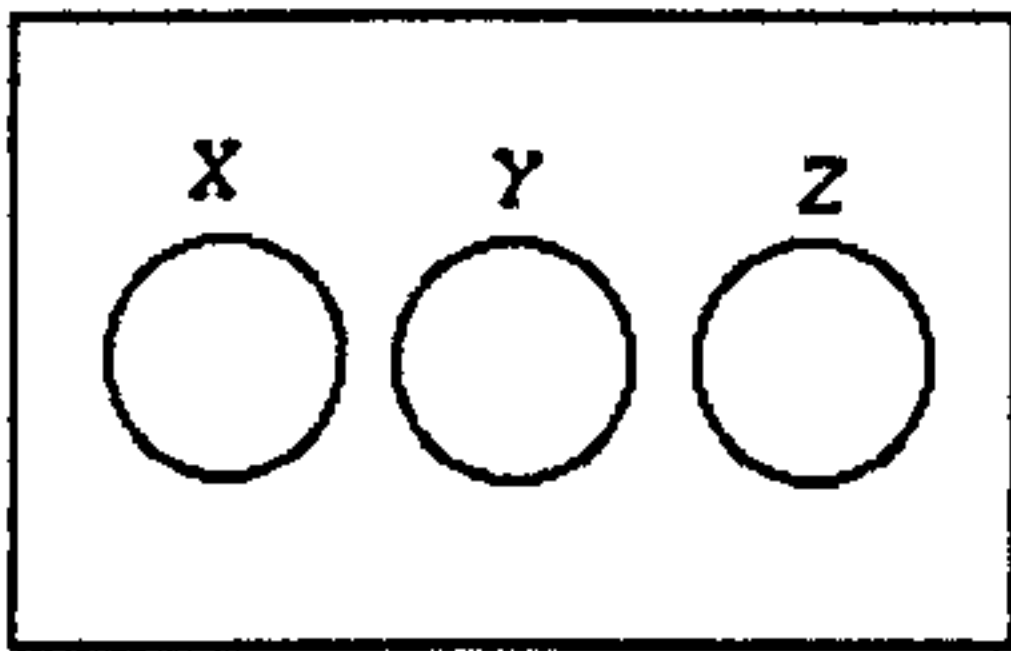
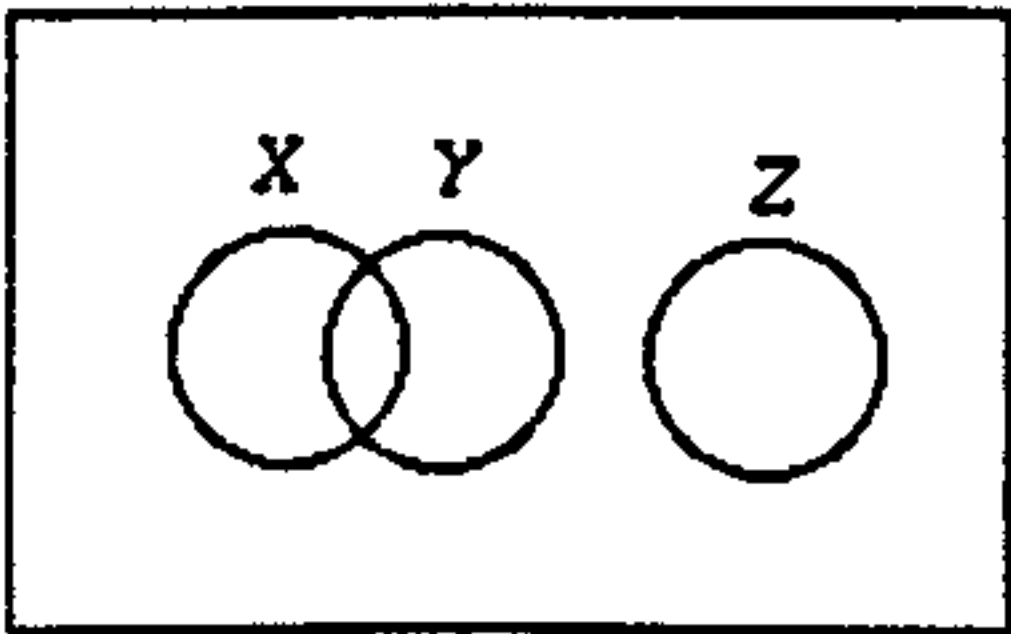


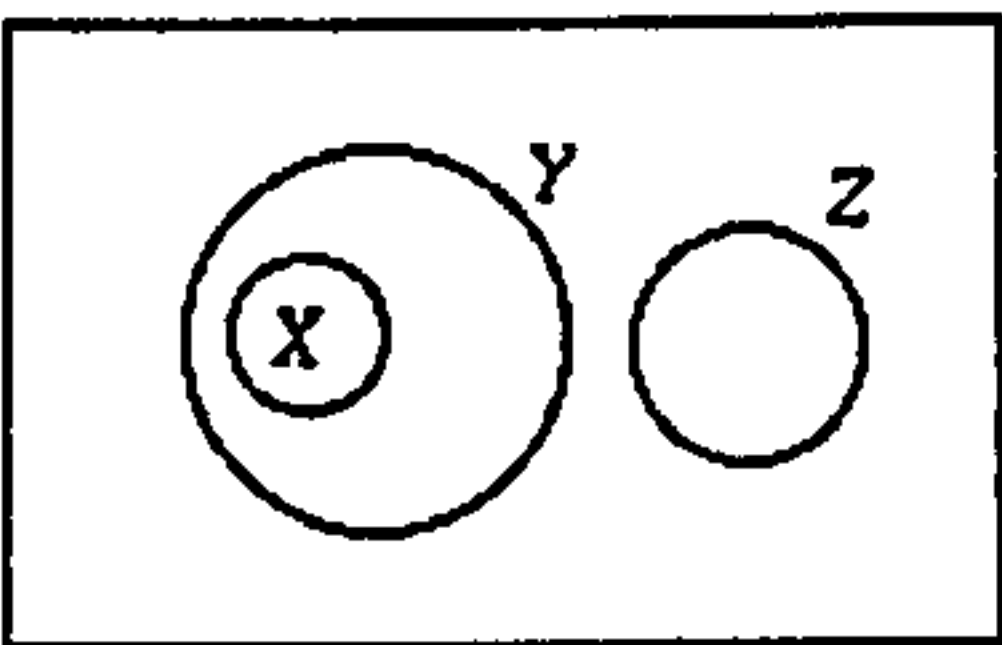
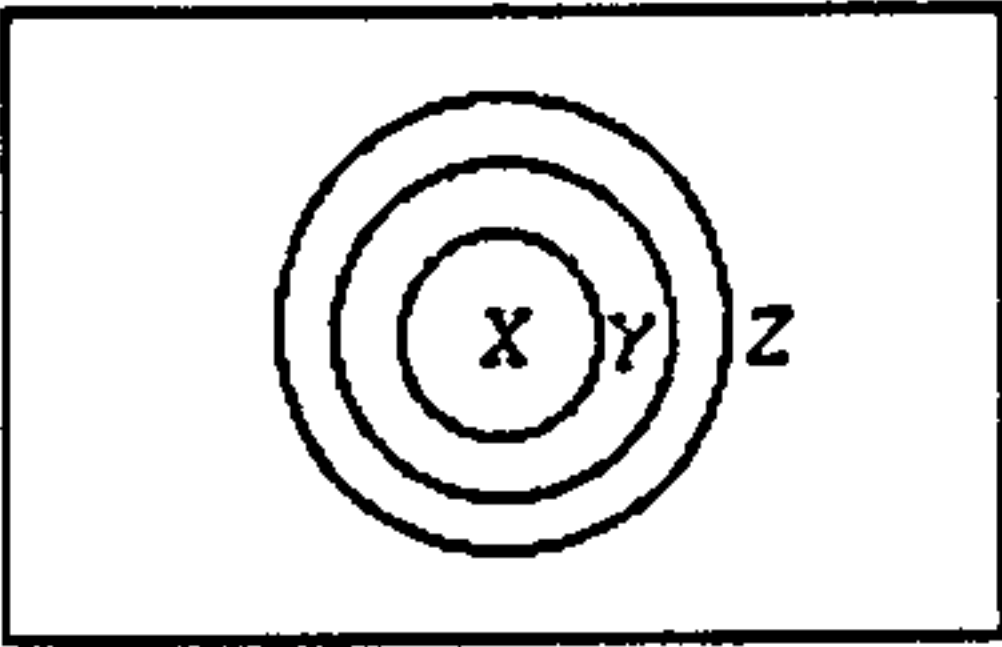
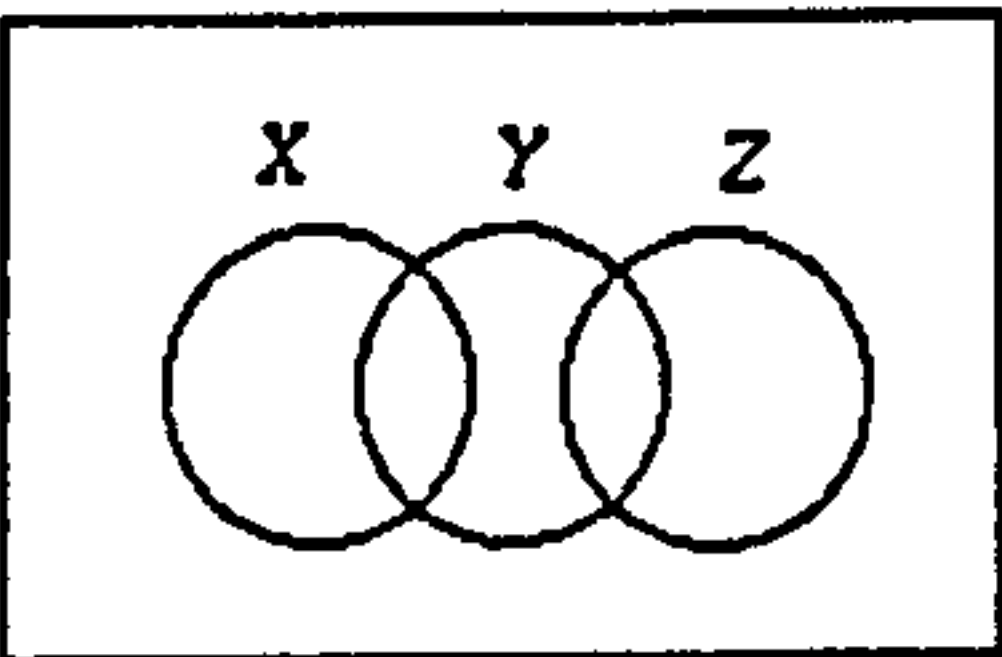
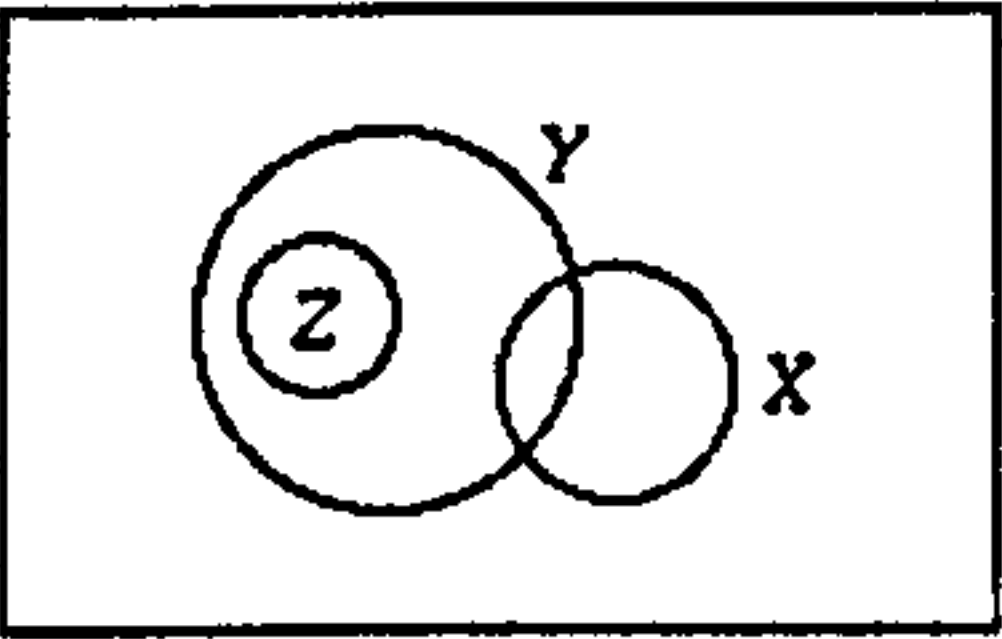
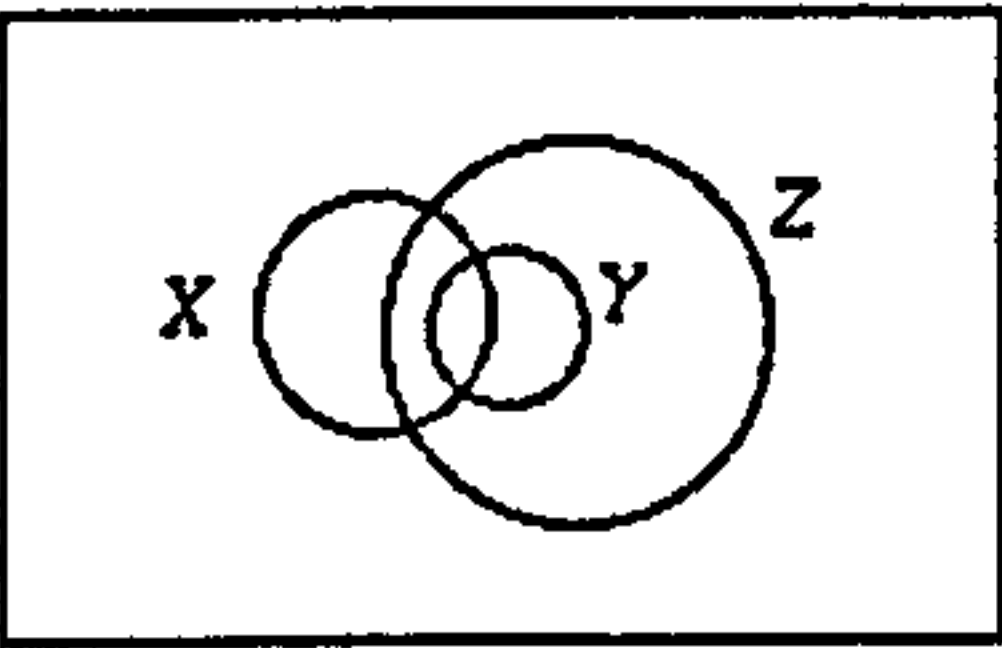
This example shows No As are Bs in the first part of the diagram, then All Bs are Cs in the second part. A natural language example of this might be No Architects are Builders, All Builders are Carpenters. Registration of the B circles leaves the one nonintersected region C, B & \_A. The A circle cannot intersect with the B circle since the first premise prohibits it. When the B term is removed a region called C,\_A is derived. The region is not labeled by an end term so can have only an existential conclusion, that is, Some Cs are not As.

The first stage in assessing the expressivity of Edinburgh Euler is to count the number of states that the system legally allows. In its most general form, the Euler system is shown with three circles that can be arranged in any possible pattern. Each of the regions in those patterns are either shaded or unshaded. The fourteen diagrams shown in Table 8 illustrate the significantly different ways in which three circles can intersect. The circle pictures are described with text and with set notation.

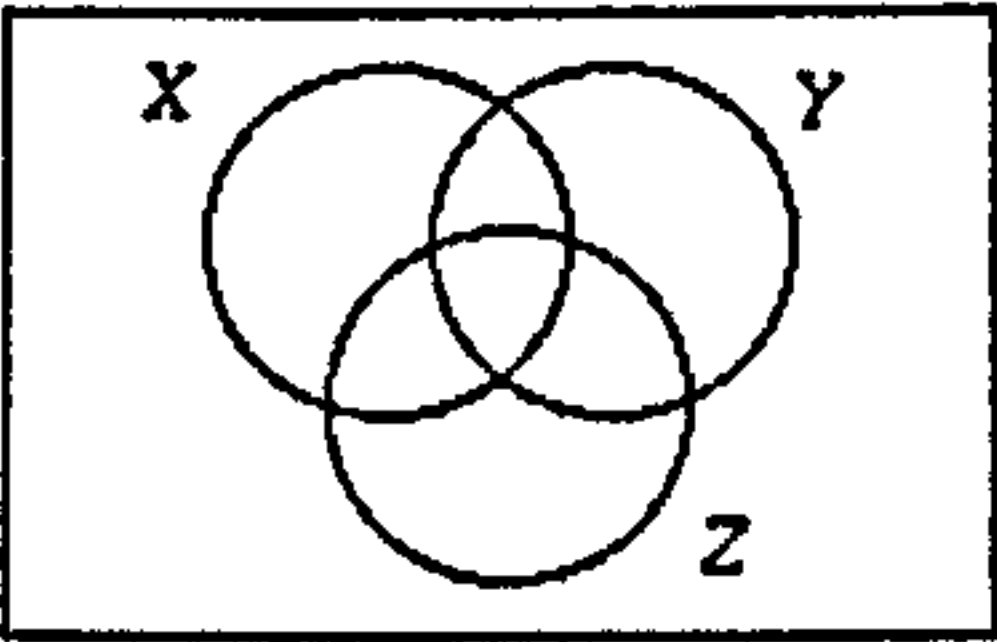
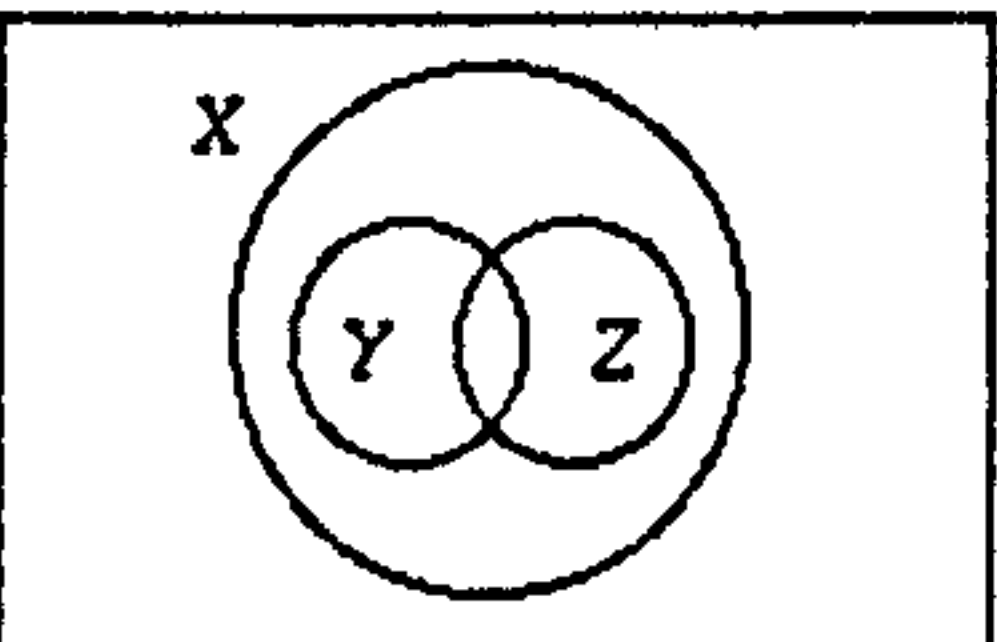
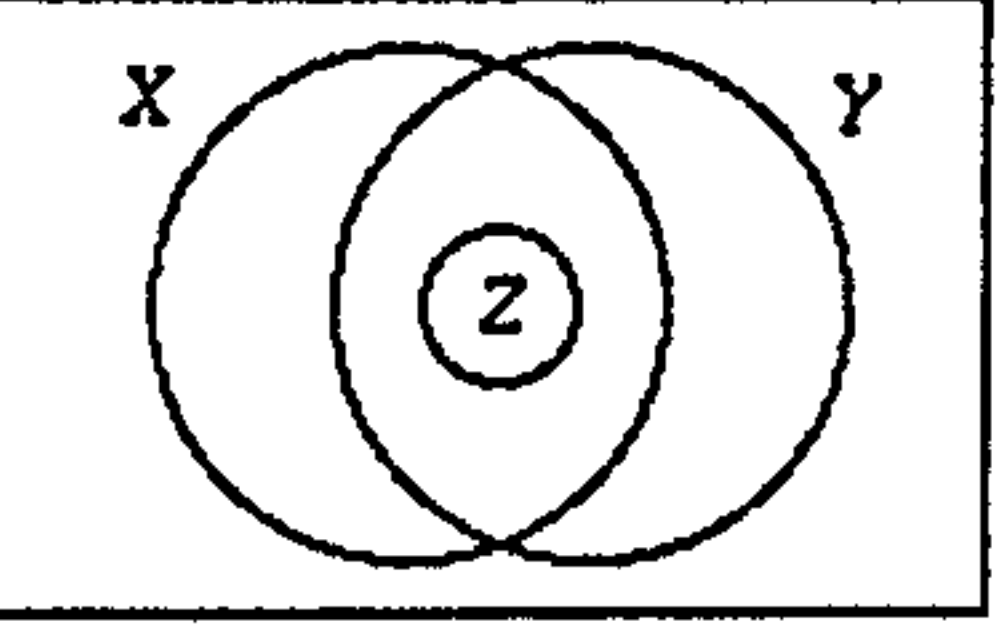
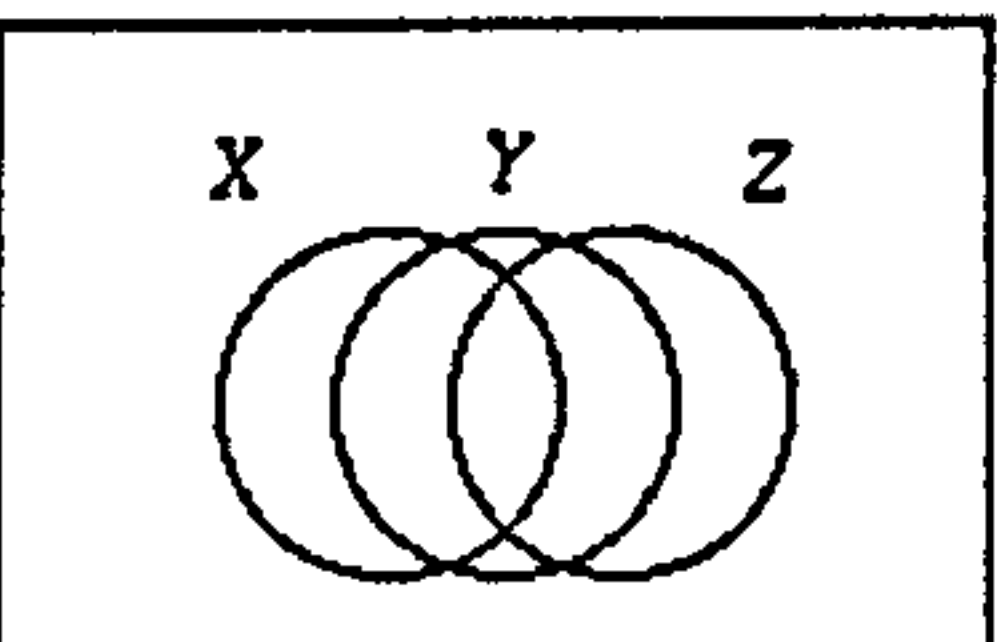
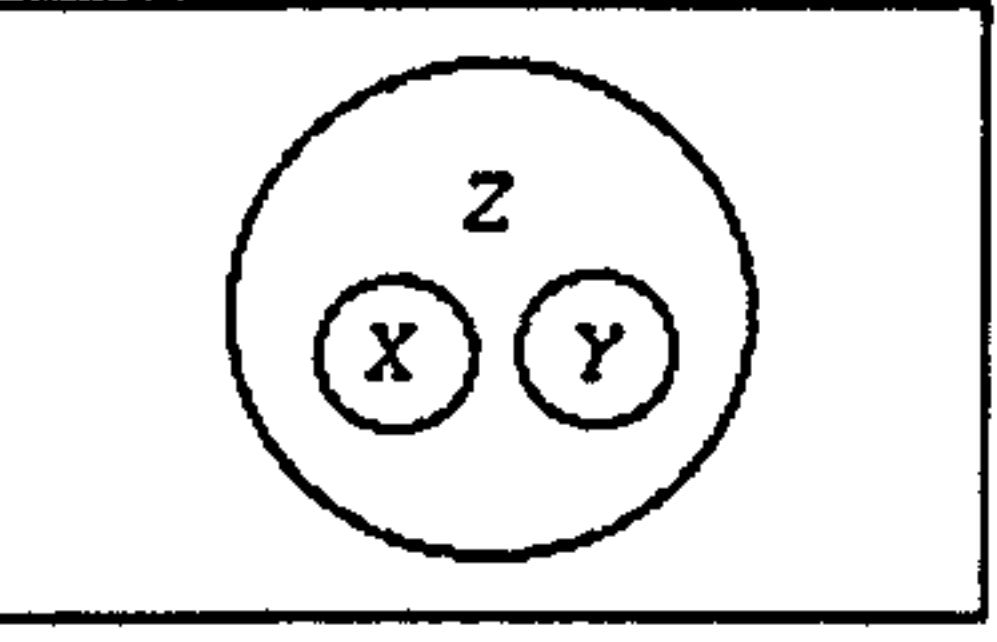
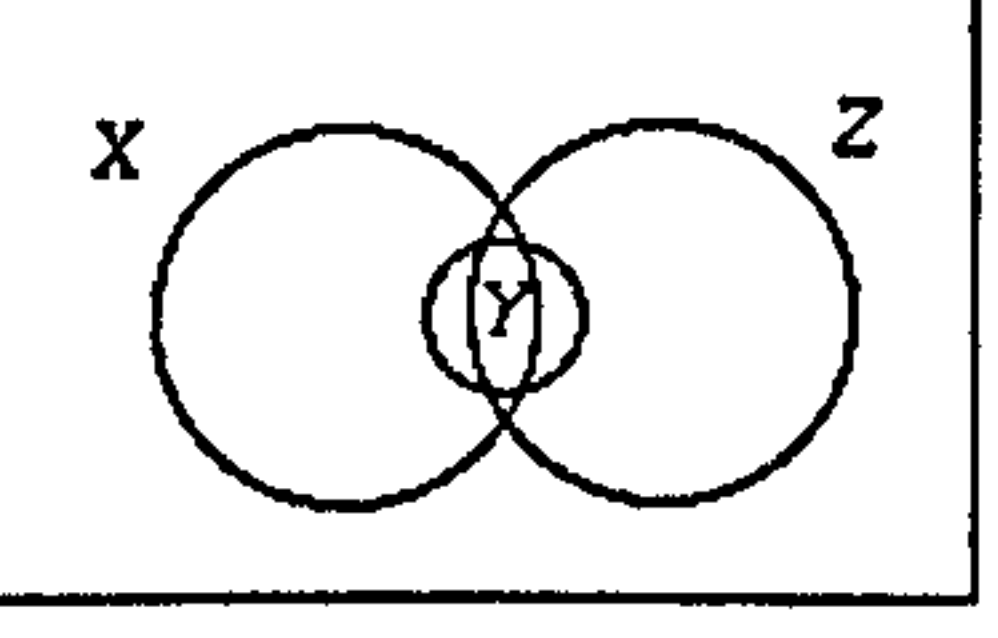
Each diagram has permutations of its own. The three circles are attached with the labels X, Y and Z. These labels indicate the names of predicates that describe the occupants of each set. The labels can be switched between circles in each diagram. Sometimes this does not effect the meaning of the model (e.g., diagram 1). In other cases, switching the labels increases the number of models that the diagram can represent. Each diagram has a number of identifiable regions. These are the result of the intersection of circles. Each region is a possible location for shading. The number of models for each diagram, is multiplied with the number of annotation opportunities, to produce the total number of states for that diagram. Adding this calculation for each of the fourteen diagrams produces the total number of states in this general form of the Euler system.

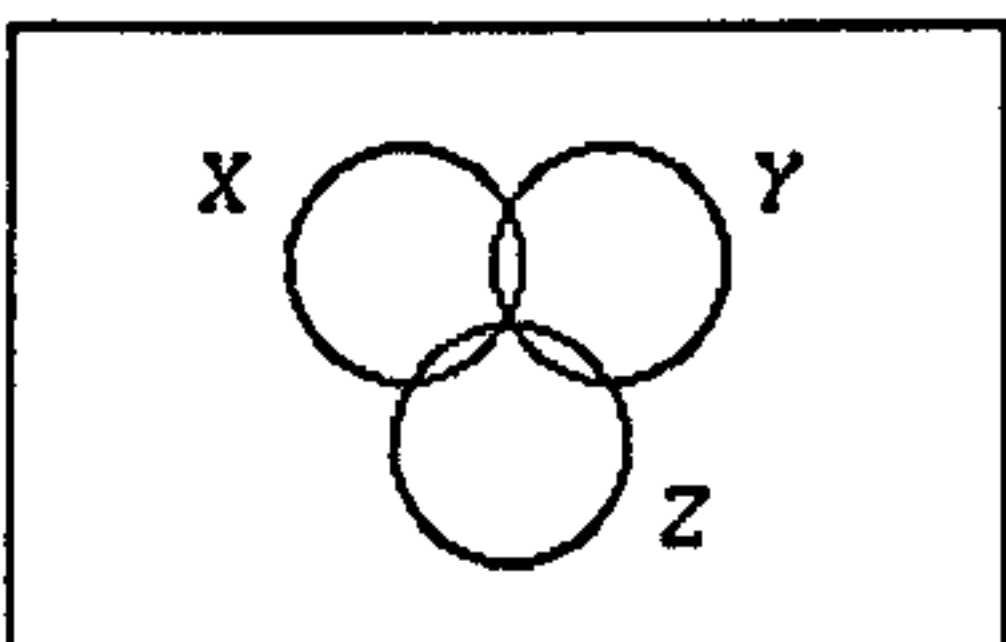
Table 8: Possible characteristic diagrams in Euler.

Diagram	Text description	Descriptions	M	R	States
1. 	<i>There are three circles without intersection. In this case there can only be one possible model.</i>	$X \cap Y = \emptyset.$ $Y \cap Z = \emptyset.$ $Z \cap X = \emptyset.$	1	3	$2^3$
2. 	<i>There are two intersecting circles and one circle separate from the others. In this case, the Z circle could be either A, B or C and so there are three possible models.</i>	$X \cap Y \neq \emptyset.$ $Y \cap Z = \emptyset.$ $Z \cap X = \emptyset.$	3	4	$3 \cdot 2^4$

3.		<p>When one circle is entirely contained within the other and the third is separate. In this case, the Z circle can be identified with either A, B or C terms, and for each arrangement of Z, there are two possible arrangements of Y and X.</p>	$Y \cap X = X.$ $X \cap Z = \emptyset.$ $Y \cap Z = \emptyset.$	6	3	$6 \cdot 2^3$
4.		<p>One circle is contained within another and in turn that circle is contained within a third. Again, Z can be assigned to either A, B or C, and for each of those arrangements there are two ways to assign X and Y, yielding 6 independent models.</p>	$Y \cap X = X.$ $Z \cap Y = Y$ $Z \cap X = X.$	6	3	$6 \cdot 2^3$
5.		<p>Each of three circles intersect but with one in the middle connected to the two outer circles. The central circle can be either A, B or C yielding 3 models.</p>	$Z \cap Y \neq \emptyset.$ $Y \cap X \neq \emptyset.$ $X \cap Z = \emptyset.$	3	5	$3 \cdot 2^5$
6.		<p>One circle contained within another and a third intersecting just with the second. Here there are 6 independent models.</p>	$X \cap Y \neq \emptyset.$ $Y \cap Z = Z.$ $X \cap Z = \emptyset.$ $X \cap Y \cap Z \neq \emptyset.$	6	4	$6 \cdot 2^4$
7.		<p>One circle contained within another and a third overlapping both the first and second. There are 6 independent models here.</p>	$X \cap Y \neq \emptyset$ $Z \cap Y = Y$ $X \cap Z \neq \emptyset$	6	5	$6 \cdot 2^5$



8.		<i>All three intersecting with common region. In this arrangement there is no way to assign labels in any other way than that shown, therefore, there is only one model.</i>	$X \cap Y \neq \emptyset$ $Y \cap Z \neq \emptyset$ $X \cap Z \neq \emptyset$ $X \cap Y \cap Z \neq \emptyset$	1	7	$2^7$
9		<i>One intersects another, both within a third circle. The outer circle could be assigned to either A, B or C and so there are 3 models here.</i>	$X \cap Y = Y$ $X \cap Z = Z$ $Y \cap Z \neq \emptyset$	3	4	$3 \cdot 2^4$
10.		<i>One circle in the intersection of another two circles. Again, there are three possible assignments of the central circle.</i>	$X \cap Y \neq \emptyset$ $Y \cap Z = Z$ $X \cap Z = Z$	3	4	$3 \cdot 2^4$
11.		<i>One circle obscures the intersection of two others and intersects with external space yielding 7 regions. The central region can be mapped to either of A, B or C and so there are 3 possible models.</i>	$X \cap Y \neq \emptyset$ $Y \cap Z \neq \emptyset$ $X \cap Z \neq \emptyset$ $X \cap Y \cap Z \neq \emptyset$	3	7	$3 \cdot 2^7$
12.		<i>Two non-intersected circles contained within a third. In this case there are three possible assignments which can be made to Z, and so three models.</i>	$X \cap Y = \emptyset$ $Z \cap X = X$ $Z \cap Y = Y$ $X \cap Y \cap Z = \emptyset$	3	3	$3 \cdot 2^3$
13.		<i>One circle intersects with two others - leaving 7 remaining regions and 3 independent models.</i>	$X \cap Y \neq \emptyset$ $Y \cap Z \neq \emptyset$ $X \cap Z \neq \emptyset$ $X \cap Y \cap Z \neq \emptyset$	3	7	$3 \cdot 2^7$

14.		<i>Three circles intersect each with two others but with no common intersected region exists. This diagram leaves us with 1 model and 6 separate regions.</i>	$X \cap Y \neq \emptyset$ $Y \cap Z \neq \emptyset$ $X \cap Z \neq \emptyset$ $X \cap Y \cap Z = \emptyset$	1	6	$2^6$
-----	---	---	--	---	---	-------

---

Total states	1600
--------------	------

---

*Note:* M = Models, R = Regions. Left column refers to a particular circle arrangements. Shading is not shown

The first two diagrams in the table illustrate the method used to count possible states in Euler. The first diagram in Table 8 has three completely separated circles labeled X, Y and Z. If the order of the labels were changed there would be no consequent effect on the meaning of the diagram. This is another way of saying that the diagram supports only one model. In counting regions, the domain of discourse is not included as a normal region since it is the region in which all other regions exist, and must therefore be present for all diagrams. In summing regions for this first diagram, there are three circles represented by R in Table 8. Each circle can be either shaded or not shaded, so the number of states sums to  $2^3 = 8$ , for this first diagram.

The second diagram supports more models than the first since the possibilities for allocating three term labels, A, B and C to each of the circles, generate diagrams with different semantic content. Either of the terms may be unified with the Z circle in the picture and the other two terms would then intersect as X and Y in the picture. There are three such models corresponding to each of the three terms (A, B and C), unifying with the Z circle. There are therefore three models for diagram 2. This diagram supports four distinguishable regions and the number of states which may be achieved from the diagram is the product of all regions and models. In this case the sum of all regions is  $3 \cdot 2^4$  possible states.

This method of summing the possible states of Euler circle system eventually adds to 1600 possibilities.

4.4.1 Reducing the expressivity of Euler.

Forming a conclusion with the Euler system depends on the combination of all term circles with as many intersections as are possible. This conclusion diagram is maximally intersected and is the only combination of the premise diagrams that can determine a legitimate conclusion to the syllogism.

In most syllogisms this final diagram is not the only one consistent with the problem premises. This is easily illustrated with a single premise. The diagram in Euler, that represents



the premise Some As are Bs, consists of two intersecting circles with shading in the intersection region. The statement that Some As are Bs, however is not inconsistent with a world in which All As are Bs. A diagram that placed the A circle completely inside the B circle is still consistent with the premise. The algorithm for finding the solution requires that all the possible circle intersections are found but other diagrams are not inconsistent. These other diagrams have a relation to real situations that can be true in a possible world supported by the premises. These diagrams are also possible in the sense that the software version of the Euler system allows the user to explore them.

The method of counting the Euler circle configurations in the previous section is very similar to the method used for counting the alternative configurations in the Venn system. All permutations and configurations of the primary units in the lexicon were found and counted. This process produced a number very much lower than the Venn system. In section 4.3.2, the idea of reducing the expressivity of the Venn system was introduced. Reducing the number of permutations with any of these systems is possible by honing down the kinds of diagrams allowed. The reasoning for disallowing certain diagrams is hazardous and (as discussed already) is probably not in the spirit of the information enforcement principle. The idea of counting diagrams that are not inconsistent with a set of premises, is comparable with the method used to reduce the expressivity of the Venn diagram above.

The method for reducing the permutations for Euler begins with the correct conclusion diagram for a syllogism. All diagrams that are 'reachable' from it are counted as possible representations. In this context a diagram is 'reachable' if it represents a possible world that is consistent with the premises. The current EULER software (Stenning & Oberlander, 1995) allows manipulation of the diagrams in ways that can only produce 'reachable' and registered diagrams. The software was used to fill the entries to the following four tables.

The following figure (Figure 13) illustrates the principle of 'reachability' with the example syllogism, 'Some As are Bs, All Bs are Cs'. The figure consists of five circle diagrams. Each of these smaller diagrams show possible situations consistent with the premises.

The diagram at the top of the tree (labeled 9) is the result of combining the premises, but does not produce the largest number of intersections between all the circles. The four circle diagrams below (9), labeled (4.1), (4.2), (4.3) and (7) are other circle configurations that are possible and consistent with the premises. These numbers (4.1 *etc.* ) refer to the diagram types in Table 8 above. These diagrams are all reachable from the initially registered diagram (9). The final diagram (7) is the only one that represents the syllogism conclusion correctly for this system. The other diagrams represent situations in the world that are consistent with the premises.

The possible world *idea* of 'reachability' is physically determined by allowable movements of the circles. These constraints are inherent in the definition of circle



intersections for each of the premises. The 'B' term circle cannot be separated from the 'C' term circle, although the 'B' circle can be fully enclosed within the 'C' circle. This represents the statement that 'All Bs are Cs'. The second constraint rules that the A and B circles must always intersect. This reflects the statement that there is always at least one A that is also a B. This restriction does not limit the B circle from containing the A circle, or the A circle from containing the B circle.

These constraints allow the circles to be manipulated so the four alternative diagrams are generated. These new diagrams are consistent with the premises represented but only one has the maximum number of intersections between the three terms.

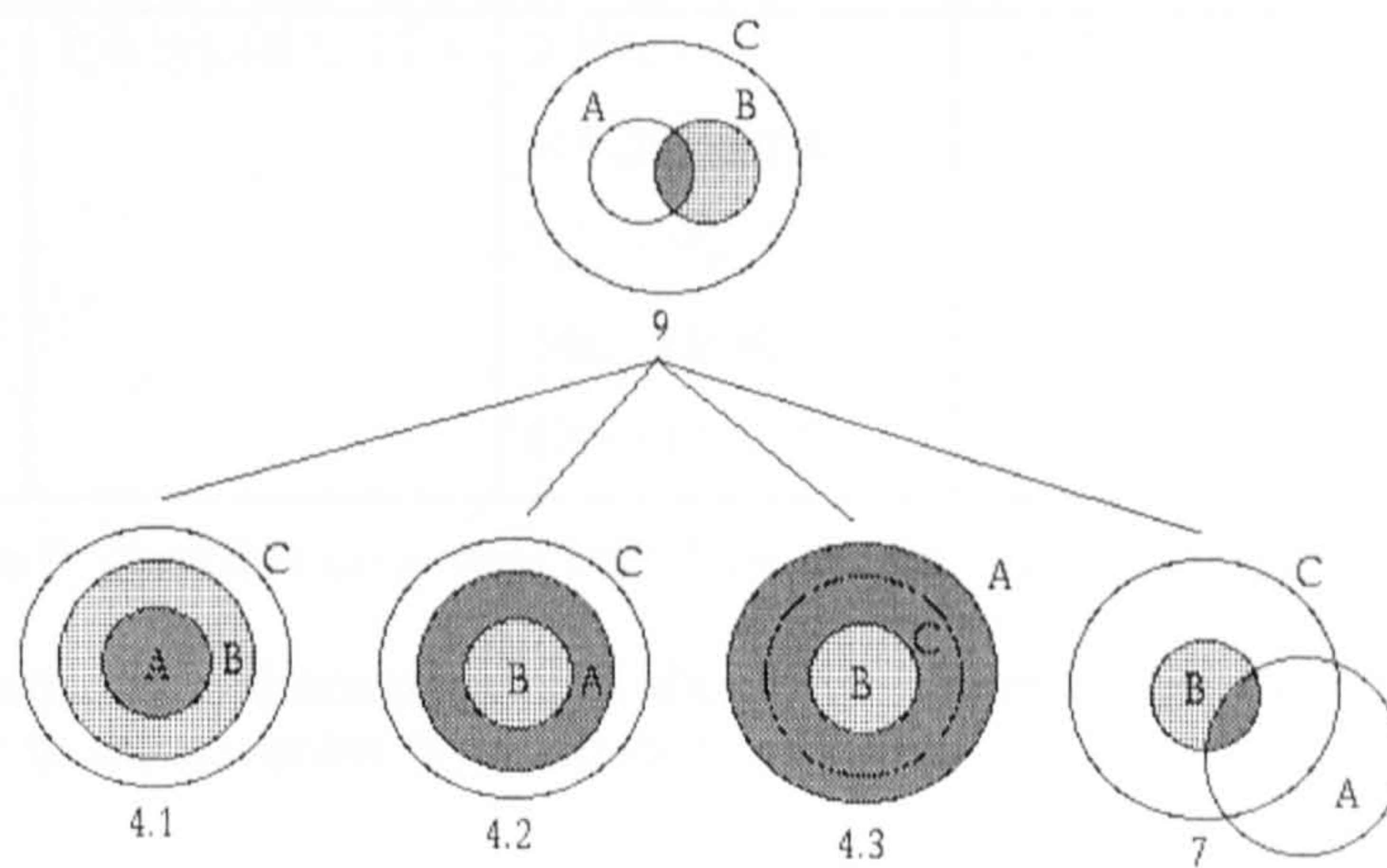


Figure 13: Counting Euler by 'reachability'.

The following tables provide all permutations for this reduced count of the Euler system. This illustrated example is included in Table 9 as a shaded cell. In all the table cells, the diagram number that represents the correct diagram for that problem is given in bold type. There are often multiple cases of the same diagram type. Variations in the labeling and shading of the same circle arrangements are counted as different diagrams. In the following tables, each multiple example is identified with the diagram type identifier (from Table 8), followed by the number of alternatives in brackets.



<i>AB-BC</i> figure.				
Premise 1.				
Premise 2.	All A-B (A)	Some A-B (I)	No A-B (E)	Some A¬B (O)
All B-C (A)	<b>4</b>	9,4(3), <b>7</b>	3, <b>6</b> ,12	3,9, <b>7</b> ,6,4(2)
Some B-C(I)	7,6,4(3),9	5, <b>8</b> ,14,6(2),12,9, 4(6),10(3),13(3) 11(3), 7(6),9(3)	2, <b>5</b> ,6(2),12,3(2)	2,3(2),4(3),5(2),6, 7(5), <b>8</b> ,9(2), 10(2),11(2),12, 13(2),14.
No B-C (E)	<b>3</b>	2(3), <b>5</b> ,6	1,3(2), <b>2</b>	1,2(2), <b>5</b> ,6,3(3), 12
Some B¬C (O)	7,4(2)6,10,9, 12,3	2,3(2), 4(3), <b>5</b> (2),6(3), 7(3),8,9(2), 10(2),11(3), 12(2),13(3),14.	<b>5</b> ,6,3,2	1,2(3),4, <b>5</b> (3), 6(4),7(4),8,9,10, 11(3),12,13(3), 14.

Table 9: Possible diagrams with EULER software in the AB-BC figure.

The shaded cell contains data also shown in the example diagram above. Entries refer to the diagram types shown in Figure 6.

<i>BA-CB</i> figure.				
Premise 1.				
Premise 2.	All B-A (A)	Some B-A (I)	No B-A (E)	Some B ¬A (O)
All C - B (A)	<b>4</b>	7,6,4(3),9,12,10	<b>3</b>	7,3,4(2),6,9,10, 12.
Some C - B (I)	9, <b>7</b> ,4(3)	5, <b>8</b> ,14,6(2),12,9, 4(6),10(3),13(3)11 (3), 7(6),9(3)	2, <b>5</b> ,6(2),12,3(2)	2,3(2),4(3),5(2),6(4), 7(5), <b>8</b> ,9(2),10(2),11 (3), 12(2),13(3),14.
No C - B (E)	<b>6</b> ,3,12	2(3), <b>5</b> ,6	1,3(2), <b>2</b>	1,2(2),3(3), <b>5</b> ,6
Some C ¬B (O)	<b>9</b> ,7,12,4(2)	2,3,4,5,6,7, <b>8</b> ,9,12	<b>5</b> ,1,6(2),3(2), 2(2)	1,2(3),3(4),5(3),6(4), 7(3), <b>8</b> ,9,10,11(3),12 ,13(3), 14.

Table 10: Possible diagrams with EULER software in the BA-CB figure.



Premise 2.	AB-CB figure.			
	Premise 1.			
	All A-B (A)	Some A-B (I)	No A-B (E)	Some A $\neg$ B (O)
All C - B (A)	4(2),9,12	7,6,4(3),9,12,10	3	3,4,6,7,10,12
Some C - B (I)	7,6,9,4(3)	5,8,14,6(2),12,9, 4(6),10(3),13(3), 11(3), 7(6),9(3)	2,5,6(2),12,3 (2)	2,3(2),4(3),5(2),6,7 (5),8,9(2), 10(2),11(2),12, 13(2),14.
No C - B (E)	3	2,5,6	1,3(2), 2	1,2(2),5,6,3(3), 12.
Some C $\neg$ B (O)	7,3,4,6,10	2,3,4,5,6,7,8,9,12	5,1,6(2),3(2), 2(2)	1,2(2),3(4),4,5(3), 6(4),7(4),8,9(2), 10,11(3),12(2), 13(3),14.

Table 11: Possible diagrams with EULER software in the AB-CB figure.

Premise 2.	BA-BC figure.			
	Premise 1.			
	All B-A	Some B-A	No B-A	Some B $\neg$ A
All B-C (A)	10,4(2)	9,4(3),7	3,6,12	3,4(2),6,7,9,10,12
Some B-C(I)	9,7,4(3)	5,8,14,6(2),12,9, 4(6),10(3),13(3), 11(3), 7(6),9(3)	2,5,6(2),12,3(2 )	2,3(2),4(3),5(2),6 (4),7(5),8,9(2),10 (2),11(3), 12(2),13(3),14.
No B-C (E)	6,3,12	2(3),5,6	1,3(2), 2	1,2(2),3(3),5,6
Some B $\neg$ C (O)	9,7,6,3,4,12	2,3(2), 4(3),5(2),6(3), 7(3),8,9(2), 10(2),11(3), 12(2),13(3),14.	5,6,3,2	1,2(4),3(4),4,5(3) ,6(4),7(4),8,9,10 (2),11(3),12,13(3),1 4.

Table 12: Possible diagrams with EULER software in the BA-BC figure.



4.5 Carroll's 'Game of Logic.'

The third system was developed by Lewis Carroll and is presented in his book, *The Game of Logic* (Carroll, 1958). The system uses a grid with four quadrants that divide the square across the horizontal and vertical axes. The upper half of the grid is a place holder for individuals that belong to the 'A' term set. The lower half of the square is a location for individuals not members of the A set.

Across the vertical divide to the left is space for members of the 'C' set and to the right individuals not in the C set. In the center of the grid is a smaller square for individuals in the 'B' term set. All regions outside this inner square are for non members of the B term set.

All possible relations between the three terms are represented by placing circular counters (tokens) into those regions. Placement of the counters indicates known existence or nonexistence and membership of the term sets.

Carroll is used to solve syllogistic problems by placing counters in the regions and on lines adjoining regions in ways that relate to claims made in the premises. Take the syllogism, Some As are not B (e.g., Some books are not novels) and All Cs are Bs (e.g., All spy stories are novels) which implies Some As are not Cs (Some books are not spy stories). This problem can be represented with Carroll as illustrated in Figure 14:

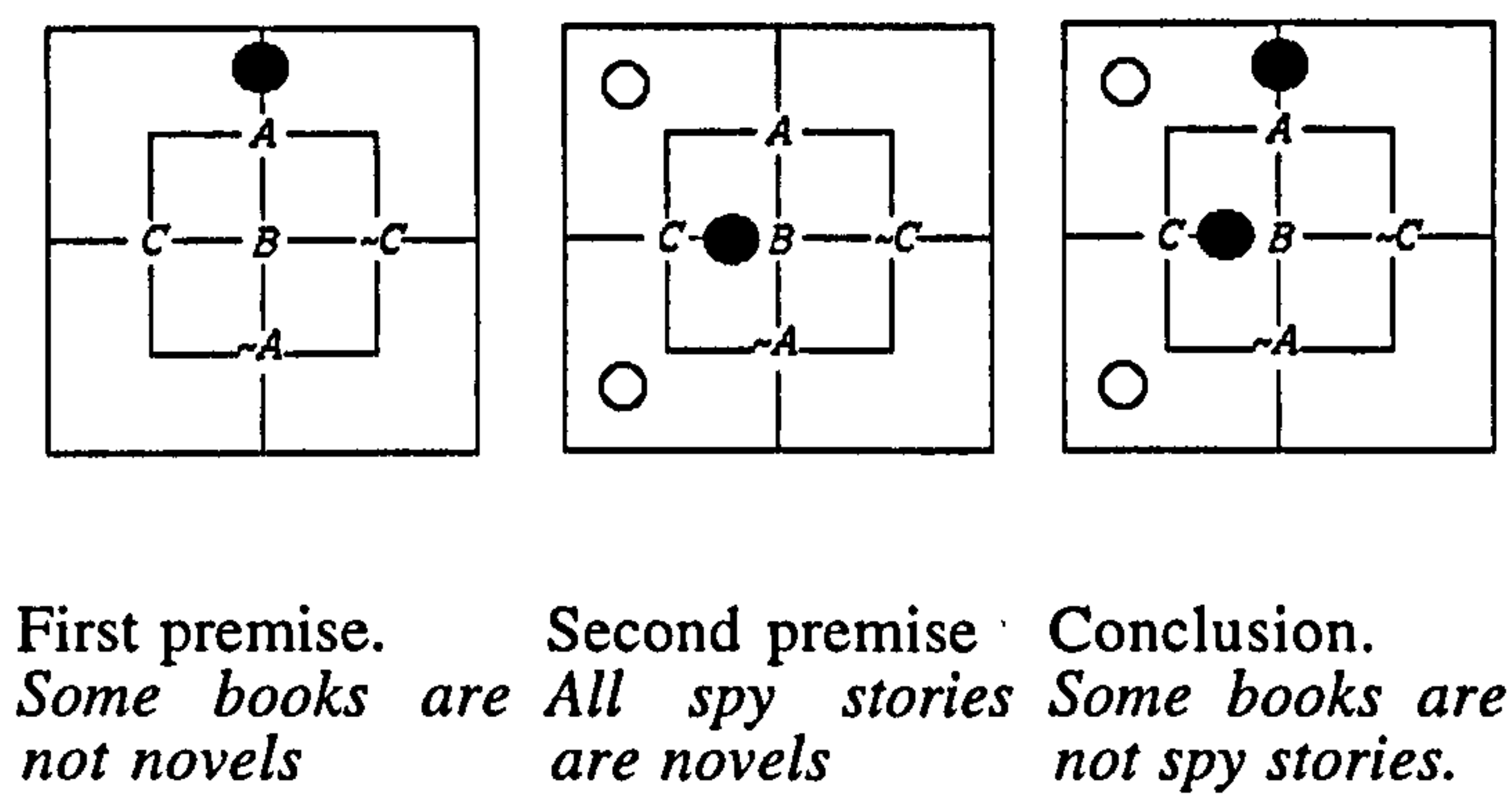


Figure 14: The Carroll system.

The Carroll system is clearly similar to the Venn system. In both systems there are twelve edges and eight regions. The regions and arcs are exactly mirrored between the two systems. Figure 15 shows the corresponding regions and arcs in each system by labeling them with the same identifier.

This figure clarifies Carroll's reflection on the improvement he had made over Venn's system. He felt that by encapsulating the domain of discourse for all other terms than the

three used in the syllogism, he had simplified the conventions. The universal set is represented by region 20 in both pictures in Figure 15. The effect of this change on the expressivity of Carroll was to increase the number of possible states, therefore reducing the specificity. According to the computational account of diagram processing earlier, this reduced specificity also reduces the benefit of the system for the learner.

The Venn system prohibits annotation of arcs that are adjoined by shaded regions. This restriction is not mirrored in the Carroll system. The complex counting in the Venn system disappears from Carroll because the same restrictions do not exist. The total number of states in the Carroll system is accordingly much higher than in Venn.

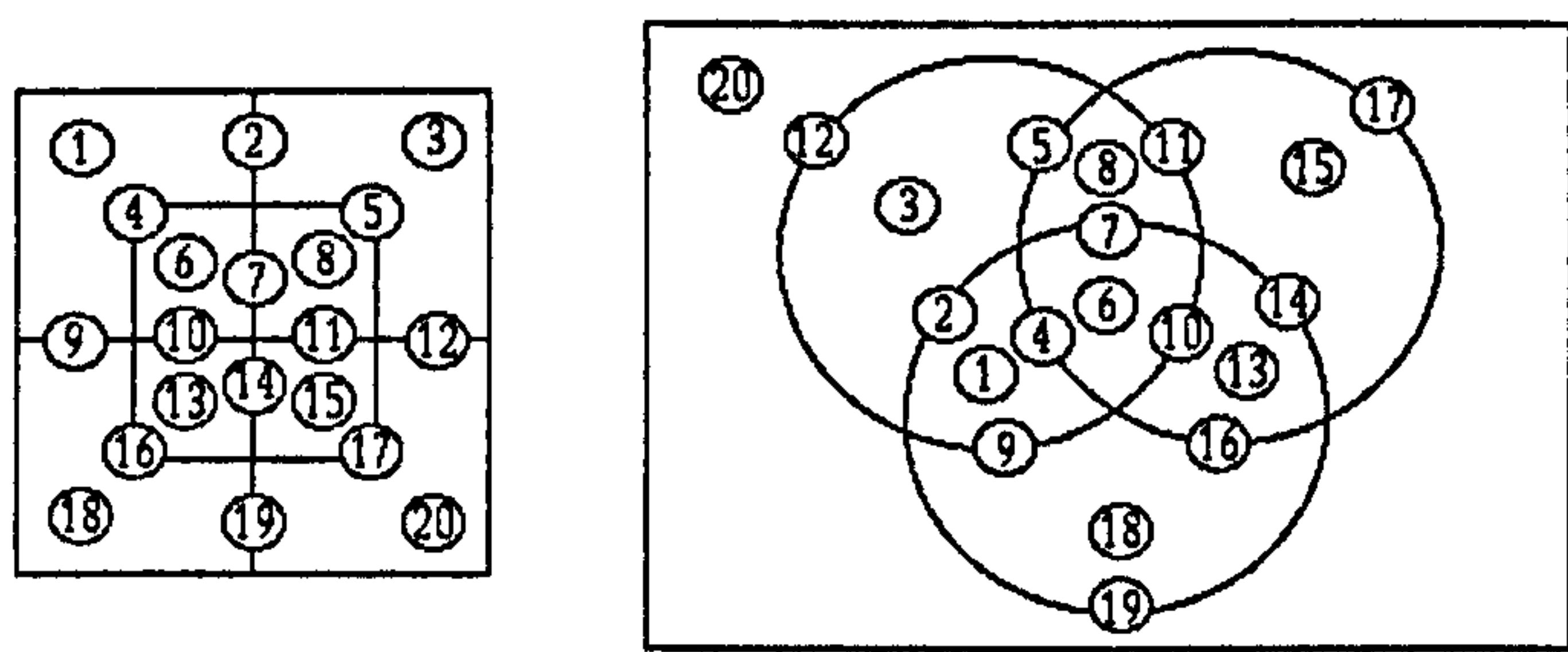


Figure 15: Counting Carroll's system.

With an unlimited number of counters of both the shaded and unshaded type, we find a similar number of states to Venn. However the restrictions that gave rise to the complex counting argument in Venn do not apply here. In Venn we could not have all combinations of states since some configurations were impossible. Representation of states not possible in Venn are now possible with Carroll. The state represented in the conclusion diagram in Figure 14 is an example of this. This is possible because Carroll uses counters instead of the shading. A region with no members is merely indicated to have no members by virtue of the shaded counter it contains. Unlike the complete shading of regions in Venn, the shaded counter in Carroll does not preclude unshaded counters from occupying the adjoining arcs. The implication of this lack of restriction in Carroll, is that the full  $3^{20}$  or 3,486,784,401 states are allowable.

#### 4.5.1 Reducing the expressivity of Carroll

As with the Venn and Euler systems, it is also possible to create a smaller value of expression for the Carroll system. Since learners may be aware that the syllogism has only two premises, a conclusion and three terms, it is arguable that the expressive ability for just these items should be counted. The tokens needed to communicate just these few items, are



multiplied together to form an alternative measure of expression. Since we require only two premises, each of which can only have one quantifier, we never need more than two empty counters and one shaded counter for each premise. This makes the maximum need for a syllogism having only two premises, just four empty counters and two shaded counters. Again we use the choice idea that:

$${}_xC_y = \frac{X!}{Y! \cdot (X-Y)!}$$

Although making use of the identity relation this time that;

$${}_xC_y \equiv {}_xC_{(x-y)}$$

So that the six counters are treated as 14 spaces in the system and the resultant expressivity of the system is calculated as follows.

$${}^{20}C_{14} = 38,760$$

This number is the number of states in the Carroll system, when it is assumed that there are only the number of tokens needed to display the syllogism in its normal form. The same remarks apply to this value as did to the second values in Venn and Euler. The process of introducing new constraints is rather arbitrary and endless up until the point where there are enough constraints for a single mapping with the syllogism. This would be a system with only 64 states, but with many additional constraints.

#### 4.6 Summary.

This chapter described a method for assessing graphical systems used to represent syllogistic reasoning. The method treats the number of system states as a measure of system flexibility and expressive power. The method of analysis produced values for three syllogistic reasoning systems. Each system originates from pedagogic tools developed by historically influential mathematicians involved in tutoring elementary logic to young learners (Venn, 1894; Euler, 1772; Carroll, 1958).

The evaluation method originated as a cognitive model of diagram processing (Stenning & Oberlander, 1994). The process of finding values for systems was not clearly

derivable from this cognitive model. To solve this problem, two methods of counting furnished two possible interpretations of the principle.

The first counted the product of all combinations of tokens in the system. Systems consist of tokens such as circles, counters and arcs. In each system these tokens are multiplied to create a maximum number of expressible 'statements' with the system. This number is called the *expressivity* of a system. The first method of counting yields enormous differences among the three systems. The rank order clearly puts Euler in top position (least expressive) followed by Venn in second position with Carroll some distance behind. The expression of the Euler system is 25 fold the information needed to illustrate the syllogism. The Venn system is 28 thousand fold that of Euler. The Venn system can express more than one thousand fold what Euler is able to illustrate. The Carroll system has almost two thousand fold the expressive abilities of Euler.

The assessment is consistent with comparative observations made earlier in the chapter. The Euler system is unable to represent many of the relations that Venn supports. Individuals that in Venn, can ambiguously rest between two regions, are not possible in the Euler system.

Explanation for the differences between Carroll and Venn depends on similar factors. Venn and Carroll are ostensibly very alike. Carroll observed (1958) that the main difference between the two systems was in the position taken by the universal set. There are however, the same number of regions, and the same number of arcs in Venn and Carroll. A constraint in the Venn system, that does not exist in Carroll, accounts for the large difference in expression between the two. Individuals that ambiguously belong to one of two regions in Venn are prohibited in situations where the arc is adjacent to a shaded region. The equivalent situation is represented by Carroll with unshaded counters in regions adjacent to arcs occupied by shaded counters. In Carroll there is no equivalent prohibition. This accounts for the almost two thousand fold difference between Carroll and Venn.

In some ways the Euler system is more like a non-graphical system than Venn because of the added conventions and constraints. The grammar or syntax of the system is more complex than Venn so that the complexity is hidden from analysis of the graphical part of the lexicon. The current system for analyzing specificity does not take into account this trade-off between complexity in the syntax and complexity in the semantics. Chapter 3 discusses this issue briefly, while Chapter 8 continues to propose new suggestions for future work to address this issue.

The second method of counting the systems used a more restricted set of tokens. This added restriction stems from the idea that each system has redundant expressive power that is predictable for the learner. Perhaps learners can see, for example, that there are only two premises in a syllogism. This modified method assumes surplus tokens, beyond those necessary for illustrating the syllogism, should not be counted.



Although this modified count requires fewer tokens, the total number of states for each case still results in a ratio more than one. This second count of expression reveals a different picture of the expression for each system. The second count, although interesting, is not as *natural* as the first. Table 13 summarises the results of this analysis where the systems are first counted without the restrictions (systems labeled 1) and then with the restrictions (systems labeled 2).

Table 13: Two information enforcement counts for 3 Systems.

System	Expressivity	Counting system	Specificity
Euler-1	1,600	total combination of tokens	25
Euler-2	638	no premise inconsistencies	9.96
Venn-1	1,784,593	total combinations of tokens	27884
Venn-2	346	just two premises	5.4
Carroll-1	3,486,784,401	total combination of tokens	54481006
Carroll-2	38,760	just two premises	505.6

The second method of counting the systems adds extra constraints that reduce the expression possible. A literal interpretation of the specificity principle suggests that these reduced systems ought to be better for teaching and learning than their un-cut relatives. The specificity principle explains the benefit of good systems by the limit of their expression. It follows that these reduced systems could be better than the un-cut versions.

Graphical systems are languages or systems of communication, with rules of symbol combination (syntax) and agreed interpretations of the signs (the semantics). Adding extra rules that prohibit statements that are visibly possible from the signs available, does not make the language more simple, and perhaps the opposite. The office indicator system demonstrated at the beginning of this chapter, could use just four specially designated regions, in a 20 by 20 grid. Informing the user that they should ignore the surrounding 396 squares may reduce the formal measure of expression, but still leaves a tremendously distracting and noisy system that is clearly much worse than the four squares on their own.

There are no ‘natural’ limits to this limiting process. Given enough added constraints, each system could be restricted to those states sufficient and necessary to illustrate the syllogism. All systems would then have the same optimal level of information enforcement (as measured by the current metric). Artificially reducing the number of states in this way reveals a limitation in the metric, but lends no significance to the reversal of values that compare Venn and Euler systems when counted this way.

According to the specificity principle, learners should benefit from these systems in the same rank order that describes their specificity. Groups of learners using the Euler system should be more successful than those using Venn. Similarly, groups of learners using Venn

ought to be more successful than those using Carroll. The following chapters begin to test this prediction.

Chapter 5 starts by illustrating software implemented with the interactive graphical systems illustrated in Chapters 3 and 4. Empirical studies test the predictive ability of the specificity principle in relatively authentic computer based learning situations. Differences in the functionality of the tools, described in the next chapter, produce potentially confounding factors between conditions. Supplementary materials balance the effect of these differences in later studies.





## 5 Software & Pilot Study

### *5.1 Trialing subjects and materials.*

Chapter 4 presented analyses of four graphical systems used to illustrate the syllogism. These systems all have fixed upper limits in their ability to express statements of information and may therefore be evaluated with the specificity metric. The analysis produced a rank ordering of these systems from the most to the least specific, (a) the Edinburgh-Euler system, (b) the traditional Euler system, (c) VENN, and (d) Carroll.

Other systems like TARSKI'S WORLD and the predicate calculus described in Chapter 3, cannot be counted with the metric in its current state. All these others allow a limitless number of statements. For example, there is no restriction on the length of a sentence in the predicate calculus. Although the number of symbols and the modes of relation between symbols are limited, these symbols and relations can be embedded and re-used with impunity.

There are practical limitations on the length of expressions that can be written with TARSKI'S WORLD. The software system (described below) limits the length of parsable strings to the sentence window width. This limit however, is a function of the software, and not of the abstract system it represents. The limited length of strings does not constitute a theoretical upper limit of expressible statements in TARSKI'S WORLD.

Although the specificity metric is unable to capture differences in expression between systems with unlimited upper bounds, there are still intuitive differences in specificity between those systems. The final discussion of future work in Chapter 8 begins to outline several improvements to the metric. Some of these improvements address systems with infinite expressivity.

This chapter describes a pilot study that assessed several aspects for the larger study plan reported in Chapter 6. The performance of syllogistic reasoning for untrained learners presents several errors and difficulties. Problems with syllogistic reasoning appear to be similar across cultures and ages (Johnson-Laird & Bara, 1984). The study pursued the question of learners' ability to master the subject. The topic is clearly learnable and often taught in the first year of philosophy courses as an introduction to philosophical logic. The learners in this study were younger and the study tests their aptitude for the topic. The pilot study is also a test of the teaching methods used to teach the topic. Text book descriptions of syllogistic reasoning describe the key parts of the logical form but rarely include any didactic support. The instructor in this study tries a variety of strategies for communicating the topic with learners. The results are useful in producing supporting materials for the studies to follow.



The data collected in this study tracks elementary errors and misconceptions shown by the learners. The transcripts of interactions between the learners and instructor reveal several links with the errors shown by untrained adults learning the skill.

The pilot study used three implemented software systems. Several factors influenced the choice of systems. The primary goal of the project is to provide empirical evidence for the use of the specificity metric. To this end the two implemented software systems that are currently measurable with the metric were included in the study.

The VENN and EULER systems were easy choices. The third implemented software system draws on an abstract system that is not countable with the metric. The infinite number of states possible in TARSKI'S WORLD make it incomparable with the other systems in quite the same way. Nevertheless, since it is the only uncountable and implemented system comparison with the two countable systems is still possible. Furthermore the methodology of the project suggests any data that supports improvement in understanding the field of elementary logic teaching should not be excluded. The pilot project uses all three of the implemented software systems. These systems are, (a) VENN (1991), (b) TARSKI'S WORLD (Barwise & Etchemendy, 1990) and (c) the Edinburgh Euler system (Stenning & Oberlander, 1994).

The following descriptions of each software application demonstrate the functionality of each tool. Chapter 3 describes the abstract systems that provide the inspiration for the design of these software applications. There are substantive differences in functionality between these tools. These differences demand additional materials in the following studies (Chapters 6 & 7) to off-set any unintended affects.

### 5.1.1 *Edinburgh-Euler*

Chapters 3 and 4 describe the exact meanings of diagrams used in Edinburgh-Euler and the allowable steps used to solve problems. This description focuses on the software application.

The Edinburgh-Euler software is an x-windows application written in the C programming language. It was created at the Human Communication Research Center at The University of Edinburgh, as a way to test a method of syllogistic reasoning with graphics (Stenning & Tobin, 1993) and not specifically as a teaching and learning support tool.

The application window contains several rectangular areas. Representative premise diagrams are placed into the premise diagram space in the upper half of the window. Diagrams are placed there either by selecting the relevant diagram from a graphical picking list at the top of the window, or by selecting the text version of the premise from a pop-up menu. For example, *Some A are B*, is one menu option.



There are three buttons in the lower area of the window labeled: (a) Combine, (b) Is it right? and (c) Quit. The combine button arranges the two premise diagrams into the work space window below. The combined diagram is consistent with the premises but may not illustrate all the intersections possible in the chosen problem (Stenning & Tobin, 1993; Stenning & Oberlander, 1994).

The learner must judge whether the diagram is a correct representation of the problem. When the learner produces a diagram he believes properly represents the syllogism, he can click the 'Is it right?' button. If the diagram is correct, the system will tell the learner so and will provide the conclusion to the premises. If the learner misjudges the diagram, the system responds that one or more of the regions remain nonintersected and that the diagram is incorrect.

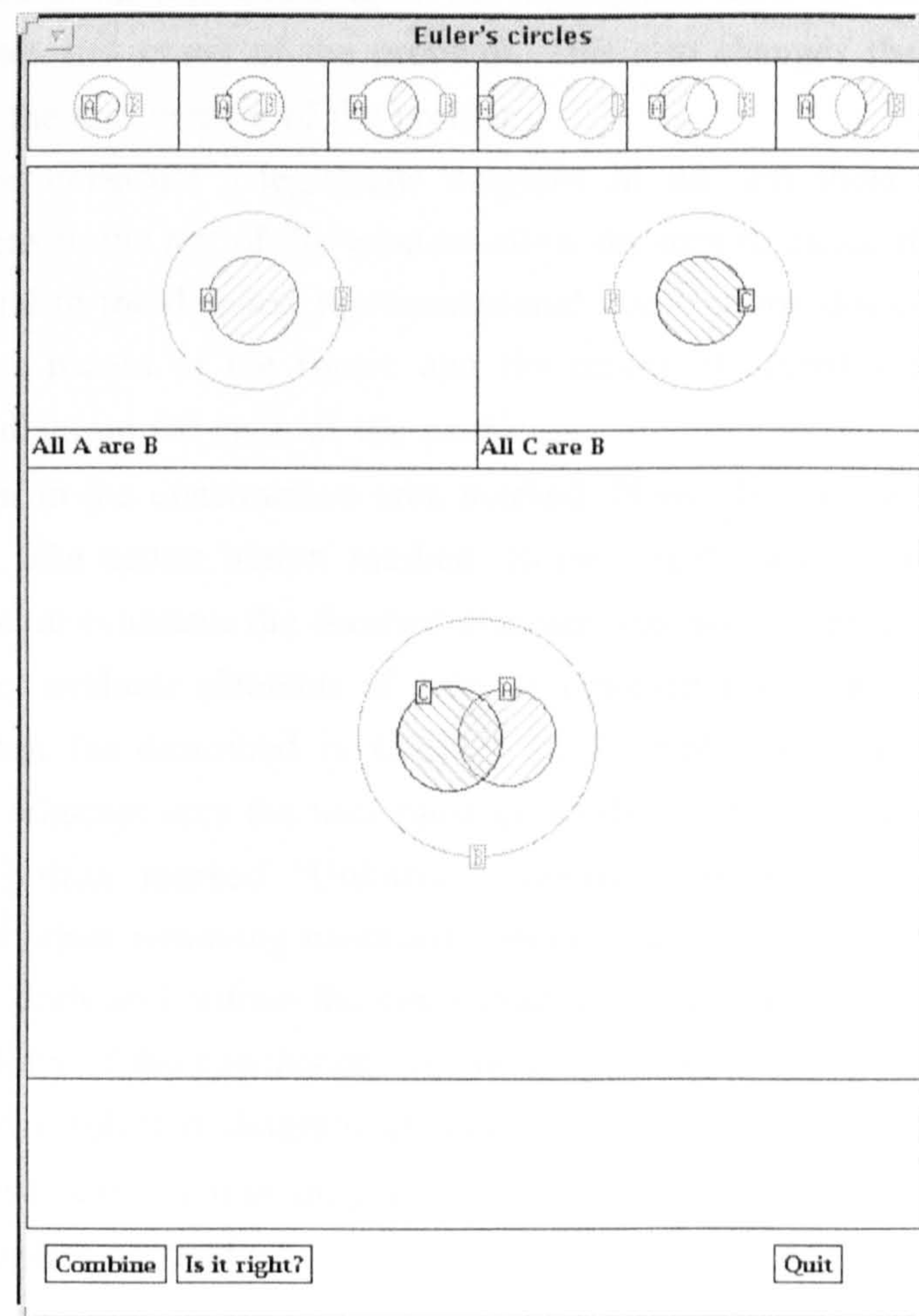


Figure 16: Edinburgh EULER software.

Note: The top row of diagrams are a visual menu for choosing premises. Lower diagrams show the individual premises and the large working space is for re-sizing and moving each circle to produce a registered diagram.



Because EULER is an x-windows application it will run most easily on a UNIX platform, although x-windows emulators are available for Macintosh and IBM compatible computers. A JAVA applet version of EULER was recently made available through the internet. The applet has most of the original functionality and since it is available with any web browser, is available on any networked computer.

### 5.1.2 Venn

The VENN software was written by Jim Moor and Marc Bedau in the philosophy department of Dartmouth college New Hampshire in the United States (Venn, 1991). The text version of the syllogism appears in the top center part of the screen and the matching figure and mood are illustrated to the left. Clicking on the highlighted parts of the mood and figure areas changes the figure and mood of the problem. This also changes the Venn diagram in the lower window and the text version of the problem.

The learner must construct a legitimate diagram in the left most side of the main screen. The three buttons to the left of the window allow the user to create this diagram. These three buttons correspond to the detailed representational conventions discussed in Chapters 3 and 4. VENN provides a record of the figure and the mood of stored syllogisms and most importantly the Venn diagram for each of the problems.

The upper button in the construction area marked 'None' fills shading into any closed region in the diagram. The center button marked 'Some' marks a star on the diagram. The illustration that the system evaluates the finished diagram and not the representation of either premise. VENN does not evaluate elements of premise representations that only occur during the annotation algorithm (as described in Chapter 3). In problems where second premises 'bounce' the 'x' from adjacent arcs the user must move the 'x' for themselves.

The lower left button marked 'Unknown' removes annotations from areas of the diagram. This is useful when removing mistaken annotations from the diagrams.

The learner may draw and redraw the circle diagrams each time evaluating the diagram and evaluating the validity of the conclusion presented by the system. At any time the learner may also have a correct solution diagram created by the software. Clicking on the 'Hide solution' button removes the solution diagram from the screen to give the learner another chance to draw the diagram correctly.



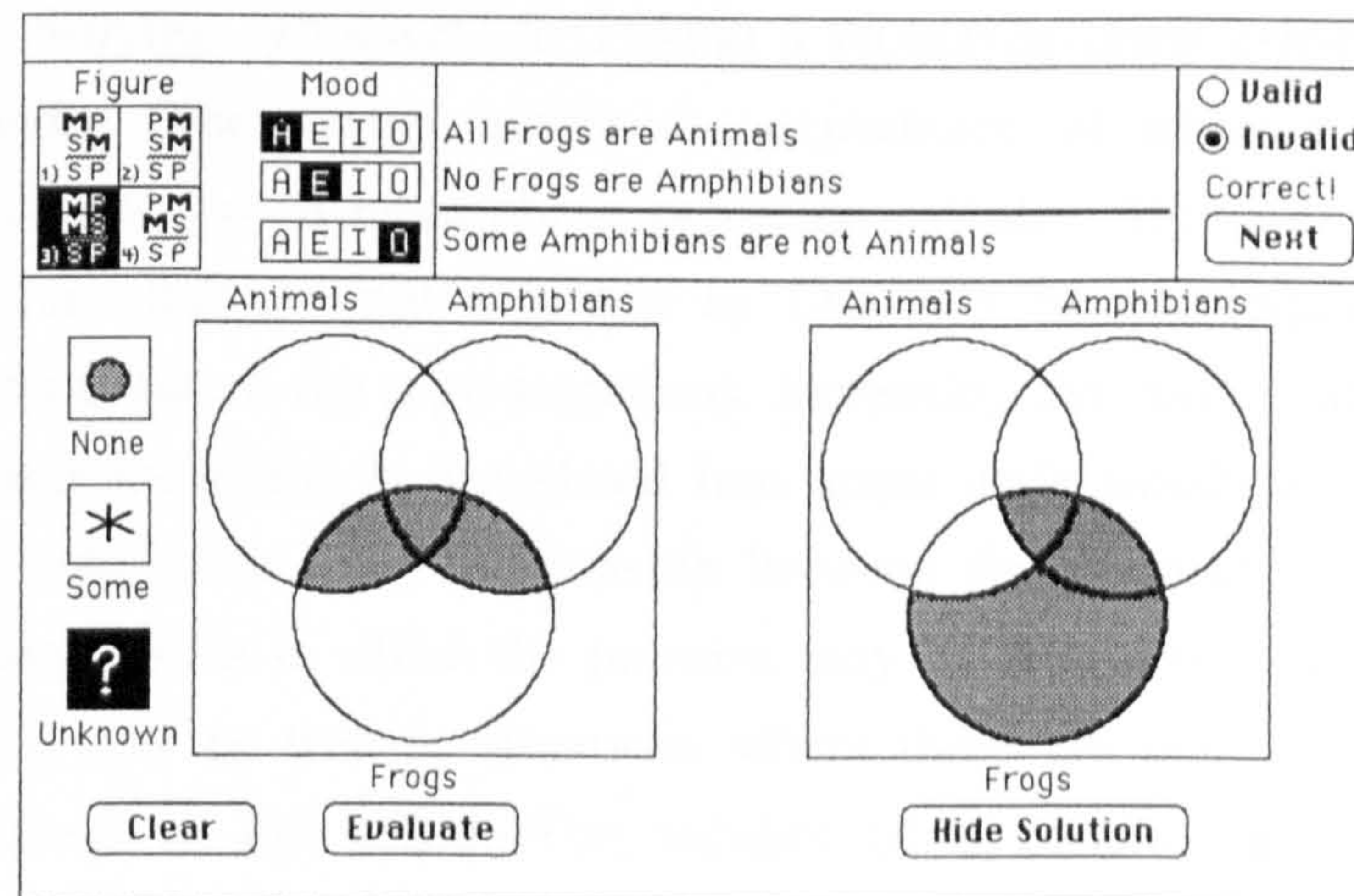


Figure 17: VENN software.

VENN unfortunately deviates from Aristotelian logic and therefore the generally understood version of the syllogism, by not accepting the inference from All to Some. As Goldson and Reeves (1992) put it, the logic does not assume the domain of interpretation is nonempty. Of the 256 possible forms of syllogism, the VENN software only treats 15 as having valid conclusions. Aristotle's interpretation of the syllogism accepts 27 valid forms. The VENN system does not allow the user to create his own syllogisms and feedback is limited to the assessment of diagrams and conclusions as either correct or incorrect.

### 5.1.3 Tarski's World

Jon Barwise and Jon Etchemendy are responsible for the development of TARSKI'S WORLD. The application is distributed with the book *The Language of First Order Logic* (Barwise & Etchemendy, 1990). The program supports an interpreted first order logic for a first order language with the addition of the equality relation. The program has three components: a world module, a sentence module, and a keyboard module. Each module appears in a separate and movable window. The world module displays certain objects such as: cubes, tetrahedra, and dodecahedra and certain relationships such as  $\text{Large}(x)$ ,  $\text{Larger}(a,b)$ , and  $\text{Leftof}(c,d)$ .

Chapter 3 describes the process of solving the syllogism with TARSKI'S WORLD. This description was developed into a tutorial to illustrate the process to learners. TARSKI'S WORLD is a semantically based system. The language used includes a limited set of recognizable objects and a limited set of real relations between these objects. Proving any statement in this kind of system demands examination of possible worlds that are true of statements offered. The process is very similar to that shown in Johnson-Laird's description of his mental model



notation (Chapter 3). Solving syllogisms in TARSKI’S WORLD involves using the objects and relations that it provides. The design makes the experience of using the software more authentic and situated as a way to learn about predicate calculus. However, problems posed with objects and relations that are not available in TARSKI’S WORLD, require the learner to translate back and forth between the representations, increasing the complexity of the task.

A pair of premises must first be translated into terms understood by the software. The learner must create a conclusion. The relationship between the premise and the situation is such that there are many worlds in which the premise may be true. For example the sentence  $\exists x (Large(x) \wedge Tet(x))$  may be true in situations where there are one, two, three, or many Large Tetrahedra present in the world. The validity of a conclusion rests on it being impossible to generate a diagram in which the premises are true and the conclusion false.

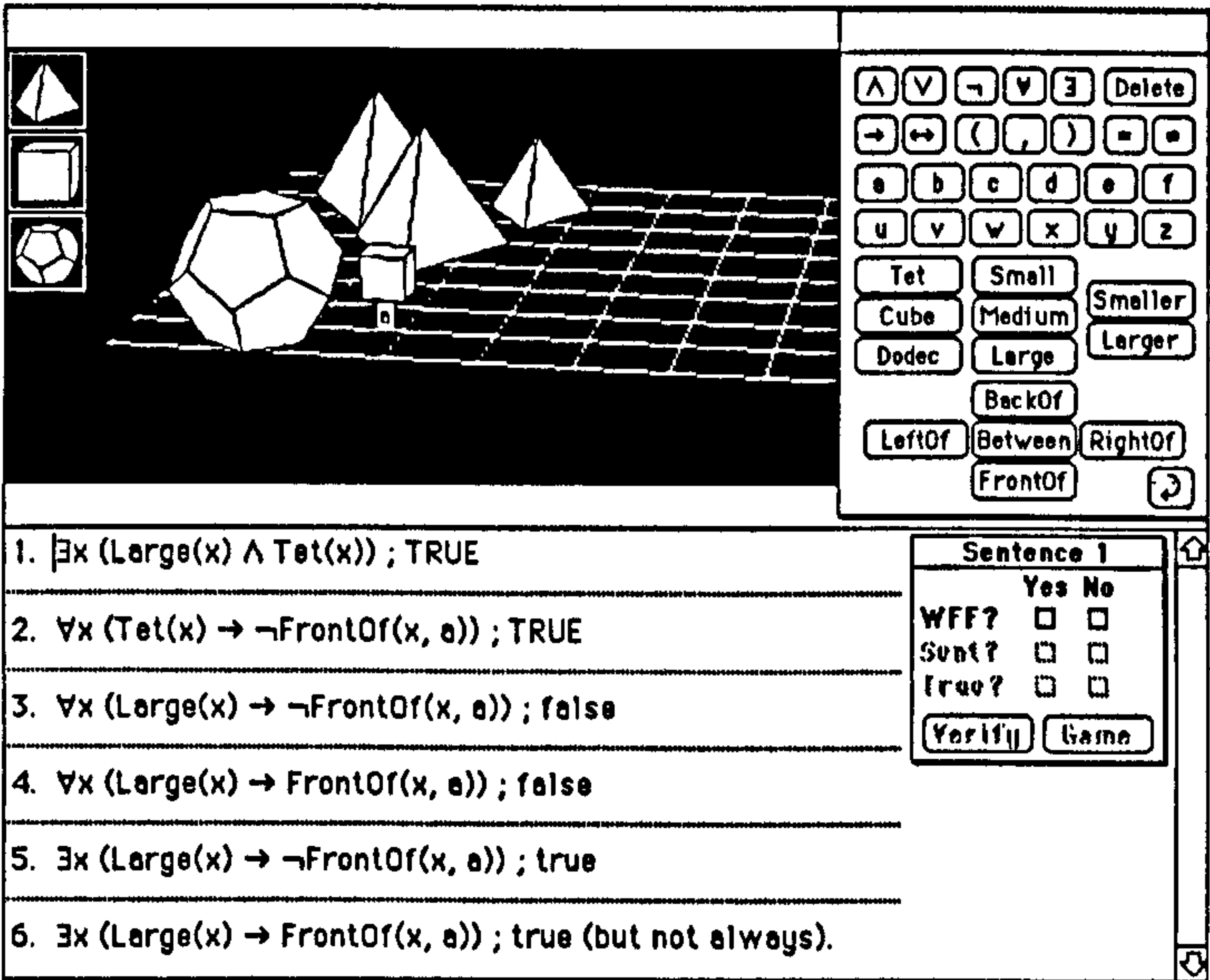


Figure 18: TARSKI’S WORLD software.

In the following example, the user receives the problem, Some Composers are Violinists, No Violinists are Drummers. They must find the conclusion by translating terms into objects and relations understood by TARSKI’S WORLD. The process may continue by assessing each of the possible forms for the conclusion in turn. In each case the learner tries to make the conclusion false by changing the diagram and maintaining the truth of the premises. Any conclusion that withstands this assessment is valid (there may be more than one).

$\exists x (Large(x) \wedge Tet(x))$	True: This Statement is true of the world built in TW. It can be checked in the verification window. (Some Large things are Tetrahedra) [Replacing Composers with large objects and Violinists with Tetrahedra and maintaining the form Some As are Bs (I), yields the first premise]
$\forall x (Tet(x) \Rightarrow \neg Frontof(x,a))$	True: This Statement is true of the world built in TW. It can be checked in the verification window. (No Tetrahedrons are infront of 'a') [Again replacing violinists with Tetrahedra and Drummers with objects in front of 'a', maintaining the form of the original second premise, No Bs are Cs (O), yields this premise in TW]
$\forall x (Large(x) \Rightarrow \neg Frontof(x,a))$	False: This Statement is FALSE of the world built in TW. It can be checked in the verification window. (No large things are infront of 'a') This statement cannot therefore be a valid conclusion from the first and second premises.
$\forall x (large(x) \Rightarrow Frontof(x,a))$	False: This Statement is FALSE of the world built in TW. It can be checked in the verification window. (All large things are infront of 'a') This statement cannot therefore be a valid conclusion from the first and second premises.
$\exists x (Large(x) \wedge \neg Frontof(x,a))$	True: This Statement is TRUE of the world built in TW. It can be checked in the verification window. (Some large things are not infront of 'a') As much as you try to change the world represented , as long as the premises remain true - this statement will also be true . It is therefore a valid conclusion.
$\exists x (Large(x) \Rightarrow Frontof(x,a))$	True (but not always). (Some large things are infront of 'a') From the diagram above this statement is also true, but if one removes the Large Dodecahedron, the premises remain true - but this statement becomes false. This statement cannot therefore be valid.

A verification window, like the one below, allows the user to check the well-formedness and truth of sentences.

Sentence 3			
	Yes	No	
WFF?	<input checked="" type="checkbox"/>	<input type="checkbox"/>	✓
Sent?	<input checked="" type="checkbox"/>	<input type="checkbox"/>	✓
True?	<input checked="" type="checkbox"/>	<input type="checkbox"/>	✓
<input type="button" value="Verify"/>		<input type="button" value="Game"/>	

TARSKI'S WORLD runs on multiple platforms including UNIX, Mac, and IBM compatible machines. It has a low memory requirement for Macintosh and can run on a range of even older Macintosh hardware such as the Mac classic. The software comes with



exercises, automatic exercise marking tools and the book, *The Language of First Order Logic* (Barwise & Etchemendy, 1990).

## 5.2 Differences among software.

The descriptions of the software illustrate systematic, pedagogic, and interactive differences among the applications.

Systematic differences reflect the underlying representations (described in Chapter 3) and the logic assumed in the software design. Both TARSKI'S WORLD and VENN both reject the implication from All to Some, however the Euler system does not. Other systematic differences between the tools are the dependent conditions for the studies that follow. There is no wish to reduce the effect of these differences.

Pedagogic differences between the tools are off-set with additional materials in empirical studies. These differences include problem sequencing (Reigeluth, 1983) feedback and instructional strategy (Dijkstra, 1997; Duffy, Lowyck & Jonnasen, 1993).

The tools described here are mostly exploratory. The tools provide a randomized or user controlled sequence of material. All the systems provide certain kinds of feedback. VENN provides visual matching feedback by changing the patterns of mood and figure at the same time as the text of a problem. VENN also allows the user to create a diagram and will then evaluate it. Quite independent of sequence, VENN also allows the user to determine if an offered conclusion is valid or invalid.

The independence of these two functions is interesting. Learners may decide the conclusion to a given problem was valid and immediately have it evaluated. There is no particular need for the user to create a diagram in this system. The user could create diagrams only when they incorrectly assess the validity of a conclusion. This behaviour did not emerge in the study.

The EULER system provides evaluative feedback for the diagram, but does not evaluate linguistic versions of the syllogism. EULER does not ask for any conclusion from the user or for the user to evaluate one offered by the system. EULER only gives the answer to the user when they have successfully created the correct diagram.

TARSKI'S WORLD has a very highly developed form of feedback. System responses to learner's suggested sentences draw on the Henkin-Hintikka game (Barwise & Etchemendy, 1991). This game takes the user's evaluation of truth and grammar of a sentence and involves the user in a dialogue. A poorly evaluated sentence triggers the game to find an inconsistency in the claim. If the user's evaluation is correct, the game will not 'find' an inconsistency, and will show the user the logic of the evaluation. The game is then a great way to check an evaluation and to learn the logic of the evaluation.

All the systems are very easy to use in an operational sense. Users showed no signs of frustration with the interface or difficulty in manipulating the diagrams. All systems, even to some extent EULER, follow a typical Macintosh interface format. Menus duplicate commands and interactions. The 'drag and drop' schema appears more in the EULER and TARSKI'S WORLD applications than in VENN.

The third category of difference is interaction. Each system provides for certain forms of interactive exploration. Creating methods to support interaction requires finding ways to represent patterns in the domain as manipulations of the interface. In syllogistic reasoning, the key structural aspects of the domain are the figure and mood. These two features describe repeatable patterns of term order and quantification. Understanding the mood and figure concepts is an early step towards understanding the full scope of all 64 problem forms. Changing the mood of any premise produces a new set of quantifiers and therefore a new problem. Exploring this transition is insightful for the learner. Changing the figure of any syllogism is simple and provides genuine insight for the user. Only VENN directly allows the user to change the mood and figure and provides a visual description of the combination of figure and mood.

Using patterns in the domain to show continuous changes in the interface allows for a special kind of interaction. The constraints of legal premise combinations in EULER makes understanding the premises nearly equal to the process of finding limits on circle movement. A circle that resists removal from within another because of limits imposed by the meaning of the representation, provides an opportunity for dynamic exploration of the constraints in the topic.

The formal descriptions of Venn do not prevent the system from similar methods of illustration. By describing the Venn system as 'registering any two well-formed diagrams,' Sun-Joo Shin (1991) suggests an interpretation much more dynamic and more like EULER, than the rather static VENN software (Venn, 1991)

TARSKI'S WORLD communicates the constraints of the syllogism through its evaluation of sentences in a simultaneously illustrated world. The constraints are looser than EULER's. A comparative constraint would be to disallow objects from worlds when their existence or relation to other objects contradicted statements made about the particular world. Adding a small cube to a world associated with a sentence claiming all cubes are large, would violate the rule. A comparable constraint for TARSKI'S WORLD could eject the small cube from the table. This would make the constraints of TARSKI'S WORLD as rigorous as EULER's.



### 5.3 The study.

This study used three mathematics students from a secondary school as case study learners. These students had not learnt any formal logic before the study although they were relatively successful mathematics students for their class.

The objectives of the study were to gauge the suitability of the subject material and the student group. The topic often forms part of a first year undergraduate course of philosophy. Learners in this pilot study were younger than the normal age of first year undergraduate students. The study was designed to determine necessary counter measures to ensure well-controlled evaluation of the specificity principle in the following studies. Text book descriptions of syllogistic reasoning rarely provide any didactic support. The study provides an opportunity to develop materials and to use the teaching and learning interactions as the basis for improvement of those materials.

Timing of the study consisted of three distinct phases over a 150 minute period. A short interactive tutorial lasted around 60 minutes. This was immediately followed by a simple test that lasted around 30 minutes. The final phase included exploratory use of the software that lasted around 60 minutes.

The teaching interaction allowed each of the subjects to become familiar with terminology and to help the experimenters develop a sense of teaching requirements for later studies. These requirements are later integrated into independent teaching booklets. Key terms were explained to the students with a short list of the terminology and some simple examples similar to those often used in introductory logic classes.

*Inference: To introduce this concept a simple example from modus ponens was given. If it is raining it is always cold. Today it is raining what else can we say about what today is like?*

*Premises: The two premises in this inference are: a) raining implies cold, and b) it is raining. And the conclusion which is in the same form is... 'It is cold.'*

*Terms: The terms are items about which the reasoning takes place. In this example the terms are: a) cold, and b) raining.*

*Truth: Truth and falsity are values which we can associate with a statement and there are two varieties for example when looking out of the window we can see if it is raining or not. If it is raining then the statement, 'It is raining' is true and if it is not raining outside then the statement is false. The statement 'raining implies cold' is true if whenever it is raining — it is also cold. This statement would have to be checked by looking out of the window too or from a database of facts about climate.*

*Validity: Validity is something we say about a sentence and its relationship with other sentences. In the example above the premises imply the conclusion and so the conclusion is valid. It doesn't matter whether it actually is raining outside for the conclusion to be valid. We say that it is valid because it has to be true if we ASSUME that the premises are true.*

*Quantifiers: There are four types of quantifiers in the syllogism which are: a) SOME, b) ALL, c) NONE, and d) SOME are NOT. A few example uses of these quantifiers: a) Some footballers wear red strips, b) All teachers are tough markers, c) Some mathematics teachers do not wear boring ties.*

The instructor presented this material in dialogue and with illustrations during the teaching interaction. The language of the topic is used quite specifically and often differently from ordinary usage. Dialogue with the learners provided a context to discuss the different uses and to equip them with the background required to tackle the example problems. Examples presented six illustrative problems that are shown in Table 14 below.

Table 14: Pilot study questions.

Syllogism	S	V	M	%E	ROD	ED	CA	CS
Some Chairs are Orange No chairs are Coffee tables No valid conclusion.	0	I				2	[I,X]	NVC
Some tables are blue All blue things are attractive Some tables are attractive	3	V	1	88	2.6	8	[I,I]	Some As are Cs
All builders are architects All builders are clever Some architects are clever	1	V	3	19	21.7	4	[A,I]	Some A-C
All astronauts are bright No Scientists are bright No Astronauts are Scientists	2	V	1	75	5.7	2	[E,E]	No As are C
Some architects are not wealthy All business people are wealthy Some architects are not bus- people	1	V	2	34	15.4	6	[O,O]	Some A not C
Some teachers are not under 18 All Children are under 18 Some teachers are not children	2	V	2	34	15.4	6	[O,O]	Some A not C

*Note: S = Score, V = Valid conclusion, M = Number of models, %E = Expected percentage of correct answers from cross cultural studies, ROD = Rank order of difficulty, ED = Number of Euler diagrams required, CA = Correct mood of conclusion and that expected by Atmosphere hypothesis, CS = Correct solution.*

The study revealed several of the biases developed in the discussion of human performance in syllogistic reasoning (Chapter 3). A rough categorization of these errors includes: (a) the atmosphere hypothesis, (b) illogical conversion and (c) the belief effect. The following transcript excerpts show examples of errors or misconceptions that are exemplars of these explanations of error.

The graphical systems created a significant area of difficulty for the learners. Learning the graphical systems clearly represents an additional load for the learner over learning the skill alone. This introduces the possibility of different kinds of error that relate to mistakes made in understanding or using the diagrams. Although introducing the graphical systems



increases the possibilities for error this does not mean that the systems make learning the skill more difficult.

In all the following transcript extracts a number of conventions increase readability. Underlining highlights spoken premises and conclusions in the dialogue. All student talk is shown in italics and labeled with S. All teacher talk is labeled T. Where more than one student is speaking in a dialogue, students are referred to by suffixes (S<sub>1</sub>, S<sub>2</sub> etc.). The transcripts do not include pauses, their duration or other non-verbal elements of the data.

*Atmosphere bias*

This excerpt from the dialogue shows development of one learner's understanding. The progression of interaction between the instructor and the two learners shows an initial conclusion that illustrates the atmosphere bias. At the end of the transcript the second learner is able to explain why the first conclusion was invalid and eventually presents the correct version.

The first statement made by the student is interesting as it does not occur as a part of the normally predicted set of errors. This conclusion is not strictly invalid but does not conform to the standard form for conclusions.

T        Here's a slightly easier one.. All chairs are orange, and, All chairs are comfortable.  
S<sub>1</sub>       All Orange chairs are comfortable  
T        No, have another go.  
S<sub>2</sub>       All orange things are comfortable  
T        Right well it is a bit more difficult than that because this [referring to orange circle] is talking about orange things and this [referring to chair circle], is talking about chairs so if we said all orange things where comfortable that would be the same as the form over here [referring to All As are Bs example] wouldn't it?  
S<sub>2</sub>       Yep.  
T        It is actually wrong even in this form. But it looks convincing though doesn't it?  
S<sub>2</sub>       *It won't be because you have not put down... If you put down All orange things are chairs and all chairs are comfortable then you could say all orange things are comfortable, but you can't at the moment because you're just stating that chair is orange and chair is comfortable but you have not said anything about other things that are orange.*

The correct solution, Some orange things are comfortable, is correct precisely because those things that are orange and comfortable are chairs. These things may have other qualities too, however it is the orangeness of the chairs that allows us to say that some orange things are comfortable. The error in S<sub>1</sub>'s conclusion is in not adhering to the convention of expressing conclusions as between the end terms only.

The move to the second statement by S<sub>2</sub> represents exactly the response predicted by the atmosphere hypothesis. The atmosphere hypothesis predicts that when neither premise is

particular, the conclusion will be universal and that when neither premise is negative the conclusion will be affirmative. In this case it will predict a universal affirmative conclusion, All orange things are comfortable.

One difficulty in using transcriptions of teaching learning interaction as corroborative evidence for mistakes and errors is that single outcomes cannot always determine a single error. In this case illogical conversion could also have predicted this outcome. The implications for bug diagnosis are well documented in other domains, for example simple two column subtraction in (Young & O'Shea, 1981; Brown & Burton, 1978). One solution to difficult diagnosis is to present problems that differentiate between competing bugs. This issue is described in the discussion at the end of Chapter 6 and in the final chapter.

### *Problems in the use of terminology*

The language of syllogistic reasoning, like most academic language, uses terminology in more specific and sometimes just differently than outside the academic context. Accelerating the learner's correct use of the terminology requires making the differences explicit. This next excerpt shows students suffering a misconception in the use of the 'Some' quantifier. The final statement from S<sub>1</sub> shows a firm commitment to his (mis)conception of the quantifier's meaning.

- T [ ]...So let's make up a premise with some quantifiers in it, we might have Some teachers wear polka dot ties. I don't know how many teachers wear ties at your school. Do they?
- S<sub>1</sub> *No not at our school , they don't wear ties at our school. [laughs].*
- T O.K. well just for the sake of the example, let's say Some teachers wear polka dot ties. Now let's imagine the staff room at the school - and that this was true. How many teachers would need to be wearing polka dot ties for this statement to be true?
- S<sub>2</sub> What, *all the time?*
- T Let's say, just on one day.
- S<sub>1</sub> *One or two.*
- T So if there were three or four or fifteen teachers wearing ties would the statement be true?
- S<sub>3</sub> *Yep*
- T And if there were just one teacher wearing a polka dot tie, would that be true then?
- S<sub>1</sub> *Umm... No, because it is then just 'a' teacher... Just one.*
- T So there is a difference between one teacher wearing a polka dot tie and some teachers wearing polka dot ties. O.K. Well in logic as long as there's one then some is true and also even if all of them are wearing polka dot ties that statement would still have been true.

### *Problems in interpreting the diagrams correctly*

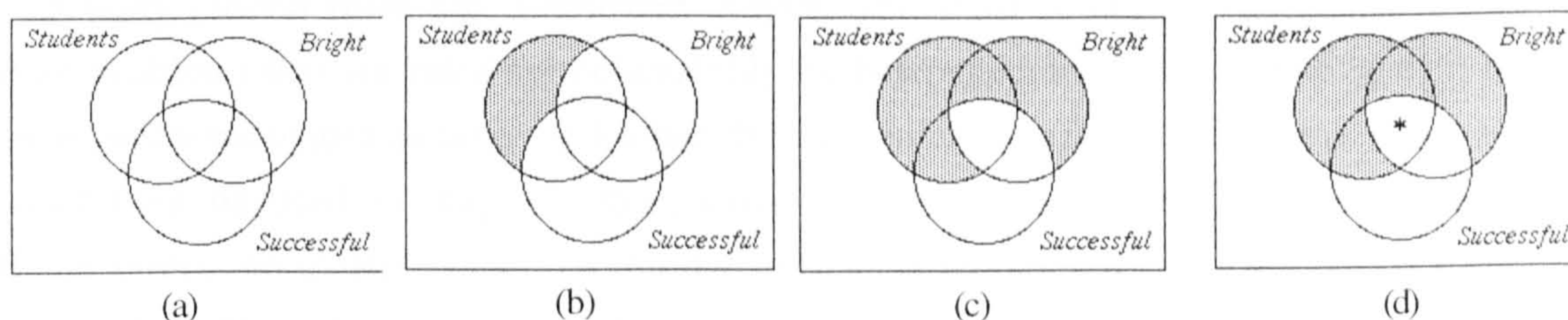
The following excerpt reveals learners' difficulties in understanding the operation of the diagram systems. This example is from the Venn system. The difficulty stems from a



stage in the process of using the diagram described in Chapter 4. When a three circle diagram illustrates a two term premise, learners are confused about the correct action to take with the third term circle. The correct action is to ignore it, since it represents the third or 'C' term. Annotation of the third circle should be left until consideration of the second premise.

The following example illustrates the error with the problem, Some As are not Bs. Although the premise only refers to A and B the learner must still remember not to add information to the diagram which makes a commitment to any knowledge about C. The dialogue should be read in conjunction with the associated diagrams (a) through (d) below

- T O.K. So in this system there are three circles, one for each term [points to diagram]. We'll do the one we got right earlier, All students are bright, All Bright people are successful. All students are successful, so we call this one students [A] this one bright [B] and this one successful [C]. So you can see that this circle is bounded by this line and they all overlap with each other.
- S *Yep*
- T And the piece in the middle is common to all three of them.
- S *Yea*
- T All students are bright. Now the convention here is to shade in the area where there couldn't possibly be any members. So if All students are bright, which part of this diagram can you shade in to say that there is nothing there? All the students are bright.
- S *The big part of students, just the large part* [points to  $\text{Stud} \wedge \neg \text{Bright} \wedge \neg \text{Succ}$ ].
- T Right O.K. this part here?
- S *The bit where there are only students and nothing else*
- T So here not there nor here nor there? [pointing to  $\text{Stud} \wedge \text{Bright} \wedge \neg \text{Succ}$ ,  $\text{Stud} \wedge \text{Bright} \wedge \text{Succ}$  and  $\text{Stud} \wedge \text{Succ} \wedge \neg \text{Bright}$ ].
- S *What already? Do you mean things which could have no members or things that we know definitely have no members?*
- T *Yep, we know definitely have no members.*
- S *Um .. just that bit then. Oh .. and the bit where there are just students and successful because all students are bright so they cant just be students and successful.*
- T Right, because we are only interested in the first premise here, so we are not really concerned with successful, only students and bright. So we could have erased this *successful* circle and it wouldn't have had any effect on the first premise.



In these diagrams the first, (a) is the empty 'set-up' diagram, (b) shows the kind of problem referred to in transcript where the student fails to see the full relation between A and B as including part of C. Diagram (c) shows the situation after two premises and (d) shows the addition of a star because of the nonempty set assumption.



In Chapter 6 this error is called 'the third term error'. The report of that study provides a full set of possible problems. Diagnosis is enabled by listing the diagram annotations to be expected for appropriate problems.

### *Belief error*

The effects of belief on syllogistic reasoning are possibly the most pervasive and easily remediated group of problems in reasoning.

- T        Some teachers are intelligent. and uh... and we'll say No teachers are under 18.  
Now.... You can see the two premises.. It's a bit like the cold and the rain thing  
[see intro on modus ponens] only its a bit more difficult. So what do you think  
follows from these two premises?
- S<sub>3</sub>        *In the second one... It depends what you define as being a teacher.*
- T        Well good point, but remember it doesn't really matter if its true about the  
world, about the school, so whatever a teacher is... the conclusion, if there is  
one, will be valid. We could replace the teacher with a tennis ball. We could say  
All tennis balls are under 18, which is probably true, but it would be a fairly  
dumb thing to say. So it doesn't actually matter if the teacher is qualified or  
not, the idea is that there is something that follows as a valid conclusion.

There are three principal ways in which beliefs can affect reasoning (Newstead, 1990):  
(a) they may distort the interpretation of the premises, (b) they may influence the deductive  
process, and bias which conclusions are reached and (c) they may be used to filter out  
unacceptable conclusions that are produced by the deductive process.

The excerpt shows learner S<sub>1</sub> exhibiting belief interference of the first kind. The subject  
is not willing to consider the process of solving the syllogism mechanically. He believes the  
contingent truth of a statement in a premise affects the validity of possible conclusions. The  
learner resists continuing the problem solving because he is concerned that the statement 'No  
teachers are under 18' may not be true.

The belief errors represent a challenge for diagnosis that may never be met without the  
kind of dialogue shown here between the instructor and learner. Learner's beliefs are  
essentially hidden from any automated system. The most promising approach would be to  
store problems that are rated for believability by independent reviewers. The system based on  
these problems would assume the learner shared similar beliefs to those in the store. Problems  
could then be used to diagnose the source, timing and incidence of belief related errors.  
Belief errors can clearly interact with other errors. A premise in the All A are B form that is  
more believable in the All B are A form would produce results similar to illogical conversion  
errors for some problems.



#### *5.4 Implications for further study.*

This chapter described the functionality and operation of three software applications for illustrating syllogistic reasoning. Differences among the applications were described as systematic, pedagogic and interactive. These are general categories intended to help think about the software. They are not meant to describe orthogonal dimensions in the designs. There are several differences between the software applications that fall into these three categories.

The support provided for exploration varies from creating possible diagrams to manipulating the underlying structure of the domain to create new problems. Feedback is provided in all systems ranging from the complex dialogue of the Henkin-Hintikka game to simple assessment of diagrams. Interaction with the learner depends on both exploration support and feedback. A slightly different form of interaction depends on the dynamic constraints of the graphical systems. Interaction based on key constraining features of the domain may be pedagogically more rewarding than interaction that is less central to the learning material. The Euler system gives learners a chance to explore the boundaries of the syllogistic form by moving circles to their allowable limits while still properly representing the premises.

The study illustrates the need to develop teaching materials that will off-set some of the major differences between the applications. The teaching interaction led to a clarification of the content required. The received meaning of terminology used in reasoning is notably different from use of the same terms in common usage. Teaching materials for the main study will include special reference to the logical meanings of these words.

Learners in the dialogue showed difficulty in producing conclusions for several reasons. Most of these reasons appear in the literature on reasoning biases. However the teaching method did not ask learners to evaluate a possible conclusion. This led to several conclusions that were only incorrect because they did not follow the usual form of the syllogism. To avoid this problem the main study will use the more limited approach of providing possible conclusions for evaluation. This strategy should avoid the problem of giving correct solutions but in unacceptable forms.

Other techniques can bring more interesting results. Johnson-Laird and Bara (1984) had students create their own solutions. Subjects find it easier to recognize valid conclusions than to generate the conclusions themselves. To evaluate learning outcomes the older method is adequate. Proportions of correct and incorrect solutions are similar in both methods (Rips, 1988).

Both the software functionality comparisons and the pilot study showed that two of the applications do not provide the learner with example problems. None of the applications have

specific structured support for the learner. For each system a specific set of steps must be taken to solve a problem. These steps must be made explicit in the teaching materials.

It will be important to control the pre and posttests to promote plausible results without the interference of learning done from pretest or from differences in difficulty of questions asked in pre and posttests. The domain offers a rich set of data on levels of difficulty for each of the forms and we are able to choose syllogisms with the same form with different terms and in different order from the pre to posttest. Finally we control the pre and posttests for belief interference by having independent adjudication for believability in each conclusion.

### *5.5 Summary.*

This chapter reported a small study that evaluated advanced high school students' aptitudes for learning syllogistic reasoning skills from interactive graphical systems. Transcriptions of interactive teaching and learning dialogues illustrate several misconceptions and errors shown by the learners. Literature describing errors made by adult populations is consistent with the learners' initial performance. Learners showed atmosphere, conversion and quantifier errors. The process of learning improved each case learner's performance. This finding provides confidence to use a similar population and learning task in the following studies.

The pilot study did not systematically collect data about the learners' use and experience with the software. A short informal focus group with the three learners suggested a high level of enthusiasm for the software. The successful use of the paper and pencil diagram systems was justification enough to introduce the software in the following studies.

The instructor's experience and the exploration of the software led to several requirements for the main study reported in Chapter 6. Training booklets use the experience of the pilot study to create general introductions to syllogistic reasoning as well as introductions to each of the software applications. The appendices of this dissertation contain the full texts of these teaching booklets.

The following chapter continues with a second phase of the empirical work, investigating the outcomes of teaching a bigger group of students ( $n = 42$ ). The results of the study reported in the next chapter represent a test of the predictions made with the specificity principle (Chapter 4).





## 6 Evaluation of the Specificity Principle

### *6.1 Motivation.*

Chapter 4 described the information enforcement metric in some detail and applied it to several syllogistic reasoning systems. Chapter 5 described a pilot study that established the selected learners could learn the material and that the software supported their learning. Chapter 5 gives a short functional description and interface screen shots for each system. This chapter describes an empirical study designed to test the effectiveness of each of the systems. The longer term goal is to test the predictive abilities of the information enforcement principle. Now that values have been attributed to the systems, it remains to test the systems in an empirical context with learners. The study presented here used 42 students from two separate streams of academic study. A mixed qualitative and quantitative methodology includes measured changes in subject's performance before and after teaching interventions. The method also analyses work-scratchings made by each learner. Results show that the information enforcement principle does not adequately predict differences among learning outcomes for all systems used. Behaviour of students whose study choices had been primarily humanities subjects was significantly different from those with mathematics interests. The responses from individual learners are collated to form the basis of an error and buggy-rule library. These errors are collated into groups under the headings of major misconceptions. Work reported in this chapter provides a first pass at a subject-phenomena generative learner model for this domain. Like the early work in subtraction (Brown & Burton, 1978; Laurillard, 1988; Young & O'Shea, 1981), these results could also be used as the basis for overlays of the possible errors for later diagnosis (Carr & Goldstein, 1977), and for other methods of learner modeling. The potential uses of this data are explored in Chapter 8.

### *6.2 Experimental design.*

This study used 42 students. All learners were drawn from classes of advanced level courses at a secondary school in Milton Keynes, England. Twenty-one of these students were studying mathematics but had no formal logic training. The remaining twenty-one belonged to a primarily humanities discipline advanced level group. Humanities disciplines ranged across Geography, English literature and History classes. Six experimental conditions were derived from the three software applications and the two discipline groups. This yields seven subjects in each condition: (a) mathematics-VENN, (b) mathematics-EULER, (c) mathematics-



TARSKI'S WORLD, (d) humanities-VENN, (e) humanities-EULER, and (f) humanities-TARSKI'S WORLD.

The study consisted of four separable parts: (a) reading the introductory tutorial and performing some simple in-text exercises. This lasted around 15 minutes, (b) a pretest that lasted about 15 minutes, (c) exploratory learning with the application software that lasted about 40 minutes and (d) the posttest that lasted about 20 minutes. The study lasted for a total of 90 minutes not including breaks between the separate parts. Short breaks between sections allowed learners to rest. The work was examined only within the experimental situation and was not intended to support main stream studies at the school. The teachers involved in the experiment were very supportive and encouraging. This undoubtedly helped to motivate students.

Subjects were drawn from an unexceptional mathematics class and a general studies class where none of the students were studying mathematics. Both groups of students came to the Open University campus at Walton Hall as part of a day visit. The subjects ate lunch and took a short tour around the campus. Both condition groups worked in small laboratories with networked computers. The machines were arranged around the outside of the rooms. An observer was present throughout the studies. Sometimes this observer was the teacher and other times the present author. For two of the conditions other staff from the Institute of Educational Technology acted as observers.

The experiment was not designed to look at collaboration between learners. Subjects in each condition did, however, work together. Learners worked in pairs discussing their approach and actions. Changes in individual scores from pre to posttest provided measures of increased performance.

Learners began the study by working through teaching materials. The first booklet is called "An introduction to Aristotelian logic". These materials compensated for possible differences between learners in their familiarity with the language of the subject area. The syllogism was presented using the usual text book method, describing language like *form*, *mood* and *term*. Elements of those materials were based on the pilot study reported in Chapter 5. The initial teaching materials were not expected to increase performance of the syllogism problems, but to establish a common level of language use for all subjects.

A pretest established the abilities of each learner prior to the use of the graphical systems. After the pretest, learners worked with a second set of teaching materials, using one of the graphical teaching systems as examples. Separate tutorials were written for each of the software systems, VENN, EULER and TARSKI'S WORLD. The purpose of these materials was to support the users when they came to use the software based on those abstract systems. These teaching materials typically demonstrate three separate problems from beginning to end. A syllogism was given at the end of each booklet that should be solved with the graphical



system. The learner should know at that point if they have figured out how to use the system and so could go back and rework the materials if needed.

Once subjects had completed these materials they were given the software application based on the abstract system studied. This period allowed time for learners to conduct experiments and practice their syllogistic reasoning skills with the software. Each tool provides some form of feedback that learners can use to evaluate their progress. Earlier remarks showed that the tools provide different levels of feedback and exploration. Paper materials were developed to compensate for those differences.

VENN produces a random list of problems that is long enough for repetitions to be unlikely. By varying the order of the terms in premises and by randomly generating the terms from the list, VENN is able to produce a sequence of problems that appear almost unlimited. In contrast, neither EULER or TARSKI'S WORLD produce example problems. A separate list of problems served as examples for learners using these applications.

Ultimately it was not possible to off-set all the possible side-effects from differences among the tools. It was not possible, for example, to control the specific problems generated by VENN in real use. The problems that learners actually solved could have been very different from those on the test list given to subjects in the other conditions.

After exploring the syllogism with the software and working with the list of problems, learners took a posttest. The contents of this were functionally isomorphic with the pretest and so represent a test of near transfer. The measures of similarity included believability of conclusions and similarity of form. Peer adjudicators provided believability ratings for each of the problems.

Two kinds of data were collected and analyzed. Quantitative differences between pre and posttest scores provided a measure of success for each system and for each learner discipline. The posttest asked learners to produce diagram solutions for each question. These were analyzed for errors and misconceptions. Although the interpretation of this data is mostly qualitative, the frequencies and distributions of those behaviours are also recorded and described.

### *6.2.1 Learning objectives.*

The objectives of the teaching determined the nature of the test materials and the teaching materials. Performance in a test before the teaching and afterwards, measured changes in the skill of syllogistic reasoning. Both tests asked the learner to evaluate the validity of a conclusion offered in the context of two supporting premises.

Problems were presented in the standard way; each premise on a new line and a conclusion offered on a third line. The task for the learner was to determine if the conclusion followed legitimately from the premises and to indicate if it was valid or invalid.



Successfully completing this task involves several steps: (a) to interpret the form and mood of the premises correctly, (b) to understand the conventions of the system, (c) to develop diagrams for each premise in the system, (d) register premise representations, and search for contradictory models depending on the strategy required by the system they are using, (e) read conclusion from final representation, (f) to perform the task without using to software and (g) to demonstrate understanding of the concepts and relations involved.

Correct assessment of the conclusion validity implies successful completion of all these other steps. Data was also collected from the work-scratchings of learners solving the posttest questions. This provides evidence of learner's difficulties with some of the steps in the learning objectives.

Both tests grade learners on performance in the reasoning skills. There was no test of understanding the concepts and relations besides the skill performance indicators.

Current techniques for assessment of reasoning skills have tended to ask the subject for the conclusion that they think follows from the premises offered. This approach can improve the depth of data provided. This richer data holds more detailed models of the learners' understanding. The method here only reports the subjects' assessment of an offered conclusion. The technique is adequate to measure differences in value among competing graphical systems.

6.2.2 Test materials.

The pretest and posttest consisted of 15 pairs of premises with associated conclusions. Subjects were asked to judge the validity of an offered conclusion and to indicate whether they thought it valid or invalid. The posttest also asked learners to draw the correct diagram for the question using the system they had learned from the software.

Each of the pre and posttest items has a characteristic profile shown in Table 15. This table illustrates difficulty levels, numbers of models and expected correct response rates from typical learners. The first column shows the example problem in words as presented in the pretest. The second column is a description of the conclusion offered to the learner for evaluation. The letter 'V' indicates that the conclusion presented is valid and an 'I' indicates it is invalid. A 'B' or a 'U' shows the conclusion to be believable or not; this classification is derived from informal peer adjudication, as in Chapter 5. The third column headed with a 'V' indicates if there is a valid conclusion for this problem using either 'V' or 'I'. The fourth column indicates the number of models the syllogism requires (Johnson-Laird & Byrne, 1991). The fifth column indicates the percentage of correct responses in cross cultural studies, again from Johnson-Laird & Byrne (1991). The mean rank order of difficulty is presented in column six and the number of Euler circle combinations possible is shown in column seven. In the eighth column we show the correct mood of the conclusion (either A, E,



I or O), and the expected mood of the conclusion according to the atmosphere hypothesis. The last column shows the conclusion in text when there is one.

There are six syllogisms that could be answered incorrectly as a result of atmosphere errors including the single problem without a valid conclusion. There are eleven three-model problems and three one-model problems. Seven problems have believable conclusions offered that are valid. Three problems have conclusions offered that are invalid but believable and three problems have invalid and unbelievable conclusions. Only one problem has a valid and unbelievable conclusion.

Table 15 Test items used in specificity evaluation.

	B	V	M	%E	ROD	E D	CA	CS
SOME ANCHOVIES ARE BITTER NO BITTER THINGS ARE CHEWY SOME ANCHOVIES ARE NOT CHEWY	VB	V	3	38	15.2	4	[O, O]	Some A not C
NO PENCILS ARE PENS ALL PENS ARE WRITING TOOLS NO WRITING TOOLS ARE PENCILS	IU	V	3	11	24.5	2	[E, O]	Some C not A
NO RUGBY PLAYERS ARE OLD SOME OLD PEOPLE ARE FIT SOME FIT PEOPLE ARE NOT RUGBY PLAYERS	VB	V	3	20	20.7	4	[O, O]	Some C not A
ALL WARLOCKS ARE WITCHES SOME MAGICIANS ARE WARLOCKS ALL WITCHES ARE MAGICIANS	IV	V	1	86	3.2	8	[I, I]	Some A are C
ALL SAILORS ARE ATHLETIC NO SWIMMERS ARE SAILORS NO ATHLETES ARE SWIMMERS	IU	V	3	8	25.2	2	[E, O]	Some A not C
SOME WOODS ARE HARD NO HOUSES ARE WOOD SOME HARD THINGS ARE NOT HOUSES	VB	V	3	15	22.2	4	[O, O]	Some C not A
NO BULLIES ARE ARMY SOLDIERS SOME GIRLS ARE BULLIES SOME GIRLS ARE NOT ARMY SOLDIERS	VB	V	3	38	13.5	4	[O, O]	Some C not A
ALL ASTRONAUTS ARE SMART SOME TEACHERS ARE NOT SMART SOME TEACHERS ARE NOT ASTRONAUTS	IB	I	-	-	-	8	[I, X]	NVC
NO FIREMEN ARE NURSES SOME ASTRONAUTS ARE NURSES SOME FIREMEN ARE NOT ASTRONAUTS	IB	V	3	28	15.9	4	[O, O]	Some C not A
ALL BUSINESS PEOPLE ARE ASTUTE ALL BUSINESS PEOPLE ARE CLEVER ALL ASTUTE PEOPLE ARE CLEVER	IB	V	3	19	21.7	4	[A, I]	Some A are C
ALL BUSY PEOPLE ARE AGGRESSIVE SOME BUSY PEOPLE ARE CLEVER SOME AGGRESSIVES ARE CLEVER.	VB	V	1	69	8.2	8	[I, I]	Some A are C
ALL BRAIN-SURGEONS ARE AGGRESSIVE NO BRAIN SURGEONS ARE CLEVER NO AGGRESSIVE PEOPLE ARE CLEVER	IU	V	3	15	23.8	2	[E, O]	Some A not C
SOME BUSKERS ARE TALENTED ALL BUSKERS ARE CLEVER SOME TALENTED PEOPLE ARE CLEVER	VB	V	1	68	8.1	8	[I, I]	Some A are C
SOME MARATHONERS ARE JOGGERS NO MARATHONERS ARE SPRINTERS SOME JOGGERS ARE NOT SPRINTERS	VB	V	3	26	18.9	4	[O, O]	Some A not C
NO RUNNERS ARE WALKERS ALL RUNNERS ARE SPRINTERS SOME SPRINTERS ARE NOT WALKERS	VU	V	3	16	23.1	2	[E, O]	Some C not A

*Note:* B = Believability (offered conclusion can be V = valid, I = invalid, B = Believable, U = Unbelievable), V = Possible validity (V = Problem has valid conclusion, I = Problem has no valid conclusion), M = Models, %E = Expected percentage of correct responses, ROD = Rank order of difficulty, CA = Correct mood of conclusion and that predicted by atmosphere hypothesis, CS = Correct conclusion.



### 6.3 Results.

The mixed methodology produced qualitative and quantitative results. Differences between the value of each system for the learner are measured by the changes in performance shown in each group. The exploration of single test items provided a more detailed view of learners behaviour. Subjects from each of the groups worked in quite different ways.

#### 6.3.1 Quantitative results.

These results refer to the measured changes in each of the dependent conditions. Learners achieved these changes in pre to posttest scores through working with the materials and software. The two mathematics conditions with learners using VENN or EULER systems (conditions a & b) increased scores significantly from pre to posttest. In the first condition scores increased from mean of 6.2 to 12.1 and with Wilcoxin signed rank test ( $T = 3$ ,  $N = 7$ ,  $p < .05$ ). In the second condition mathematics students used the EULER system and scores increased significantly from 6.5 to 10.4 ( $T = 0$ ,  $N = 5$ ,  $p < .05$ ). Mathematics students in these two conditions showed a marked improvement in performance over the experimental period. However, learners from the same dependent group using TARSKI'S WORLD (condition c) showed no improvement at all. In two of the humanities conditions (d & e), those in the EULER and VENN dependent groups showed a significant decline in performance from mean scores of 7.3 to 5.1 ( $T = 4$ ,  $N = 5$ ,  $p < .05$ ) in condition (d), and from mean score of 5.0 to 3.6 in condition (e) ( $T = 2$ ,  $N = 5$ ,  $p < .01$ ). Differences in outcome between VENN and EULER show that the VENN system proved a significantly better support system than EULER. This was tested with the Mann-Whitney U test ( $p = 0.02$ ).

In the EULER conditions, subjects took 50% longer to achieve the improvement and to finish the posttest. This result is hard to resolve with the predictions of the information enforcement metric. Even more surprising, is the striking difference between the mathematics and humanities students in both EULER and VENN conditions (exact two-tailed  $P = .0173$ ). Mathematics students exhibited much higher learning gains over the humanities group while using either EULER or VENN. Humanities students showed universal decline in scores on posttests for solving the same syllogisms.

The students were drawn from groups with almost equally matched scores on their previous practice advanced level examinations in their own chosen disciplines. There is no reason to think the differences were a result of general intellectual abilities that could be measured by intelligence tests. The differences in outcome must depend on aspects of the populations they were drawn from. These population differences may be rooted in the learners' preferences for study or in the training they get from those choices. Learners choosing to study advanced mathematics in high school may be better suited to learn formal



reasoning skills. The mathematical training involved in advanced high school mathematics may prepare learners better for learning formal reasoning skills. The study reported in Chapter 7 looks for explanations of three differences.

There is a similarity in these findings with results found in Fung, O'Shea, Goldson, Reeves and Bornatt (1994). Their work pointed to the inability of nonmathematics students to perform as well as mathematics trained students in more complex formal reasoning skills.

There is also a similarity with Cox (1996), who studied learners' experiences with introductory logic. Classes in the Cox study received either HYPERPROOF (Barwise & Etchemendy, 1995) or the more usual syntactic approach to classroom teaching. HYPERPROOF is a graphical blocks world system for training logical proof. The Cox study (1996) shows that both conditions enabled improvement to posttest transfer domains for certain learners. For learners weak in problems benefiting from diagram use, HYPERPROOF boosted learner's performance. The syntactic version of the course appeared to degrade the same group of learner's performance. Cox splits his groups into good and poor 'model based problem solvers'. The split between the mathematics and humanities students in this study produced a similar result.

It may not be surprising that learners who choose to study mathematics find the syntactic approach appealing and useful. The stranger outcome is that poor model based problem solvers and humanities students' performance actually degrades after training with graphical representations. Possible reasons for the different performance between the two discipline groups are discussed at the end of this chapter. This discussion contributes to the design for the study reported in Chapter 7.

Table 16: Results of specificity evaluation.

Test item	Condition		
	<i>Venn</i> n = 14	<i>Euler</i> n = 14	<i>Tarski's World</i> n = 14
<i>Mathematics</i>	(a)	(b)	(c)
Pretest	6.2	6.5	8.2
Posttest	12.1	10.4	8.1
Mean Preposttest $\Delta$	5.9	3.9	0
SD	1.41	1.27	0
<i>Humanities</i>	(d)	(e)	(f)
Pretest	7.3	5.0	6.4
Posttest	5.1	3.6	6.2
Mean Preposttest $\Delta$	-2.2	-1.4	0
SD	0.78	0.21	0

Note : Values inside brackets indicate time taken to complete posttest phase in minutes.

In conditions (c) and (f) using TARSKI'S WORLD, there was no change in performance. The posttest diagrams showed that subjects had not understood how to use the diagrams to solve the problems. Very few diagrams were drawn by the subjects in these conditions, which



reflects the complexity of using TARSKI'S WORLD for this purpose. In light of this, the data from these two conditions has been discounted from the remainder of the study. Some final remarks about the reasons for difficulties with TARSKI'S WORLD are made at the end of this chapter.

The experiment showed a number of key results. According to the counting algorithm which leads to the measures of specificity and expressivity, the clear victor should have been the EULER system. However, the results have shown that VENN produced the better outcomes. The major implication of this work is that, given our interpretation and application of the specificity principle in Chapter 4, there is something wrong with using the metric to predict learning outcomes. The clearest result is in the poor performance of the humanities students compared to the mathematics students.

The third result which is strange is the unexpected decrease in learned outcome with the humanities students. It would not be so surprising that humanities students scores did not improve through teaching - lots of teaching interventions have no effect. It is more surprising that the interventions actually degraded their performance.

The next section looks at the range of problems shown by learners and at bug diagnosis from learners' drawings. After describing all the bugs produced by learners, a comparison is made between the learner groups in terms of the range of bugs which each produced. Categories of bug types and learner approaches are shown to be significantly different between mathematics and humanities students.

### 6.3.2 Qualitative results.

This section reports the results of an item analysis performed on each of the conditions (a), (b), (d) and (e). The data was produced by learners in the posttest stage of the main study. What follows is a categorization of each of the error types made by learners.

Diagrams presented in Table 17 are faithful reproductions of the diagrams made by learners in each critical element. The reproductions preserve of all meaningful topographical components including, regions, shading patterns and labeling. Shape of all closed curves are degraded to circles and completely erased diagrams are not included. In Table 17 there are seven subject's responses to a single syllogism. The complete data set could have totaled 420 diagrams, however not all subjects completed all questions and made diagrams and so exactly 274 diagrams were produced in all.

From these work-scratchings seven bug types were established from the analysis of EULER and seven from the analysis of VENN. The diagram in the top left view is the correct solution and the others numbered vertically from one through seven were made by learners. The results in Table 17 were produced from the problem, All buskers are astute people, All



buskers are clever, therefore All astute people are clever. This is an invalid conclusion and the learners should have indicated this.

Table 17: Diagrams drawn by mathematics trained Venn users.

The correct diagram.			S <sub>4</sub> No * in central region. Empty set assumption error.
S <sub>1</sub> No * in the central region. Empty set assumption error.			S <sub>5</sub> No * in central region. Empty set assumption error. Difficult to interpret extra shading.
S <sub>2</sub> This diagram is impossible to interpret.			S <sub>6</sub> No * in central region. Empty set assumption error.
S <sub>3</sub> No * in the central region. Empty set assumption error. Correctly interprets syllogism as invalid.			S <sub>7</sub> Has annotated All Cs are As, instead of All Bs are Cs.

ESE. Empty set error.	(S <sub>1</sub> , S <sub>3</sub> , S <sub>4</sub> , S <sub>6</sub> )
TOE. Premise term order error	(S <sub>7</sub> )
Path. Pathological (uninterpretable)	(S <sub>5</sub> , S <sub>2</sub> )

Note: S<sub>1-7</sub> = subject (7 in each condition), ES = Empty Set Error, TOE = Term Order Error, and Path = Pathological. This posttest syllogism was: All buskers are astute people, All buskers are clever, therefore All astute people are clever. It is isomorphic and iso-believable with, All business people are astute, *etc.*, in the pretest question illustrated in Table 15.

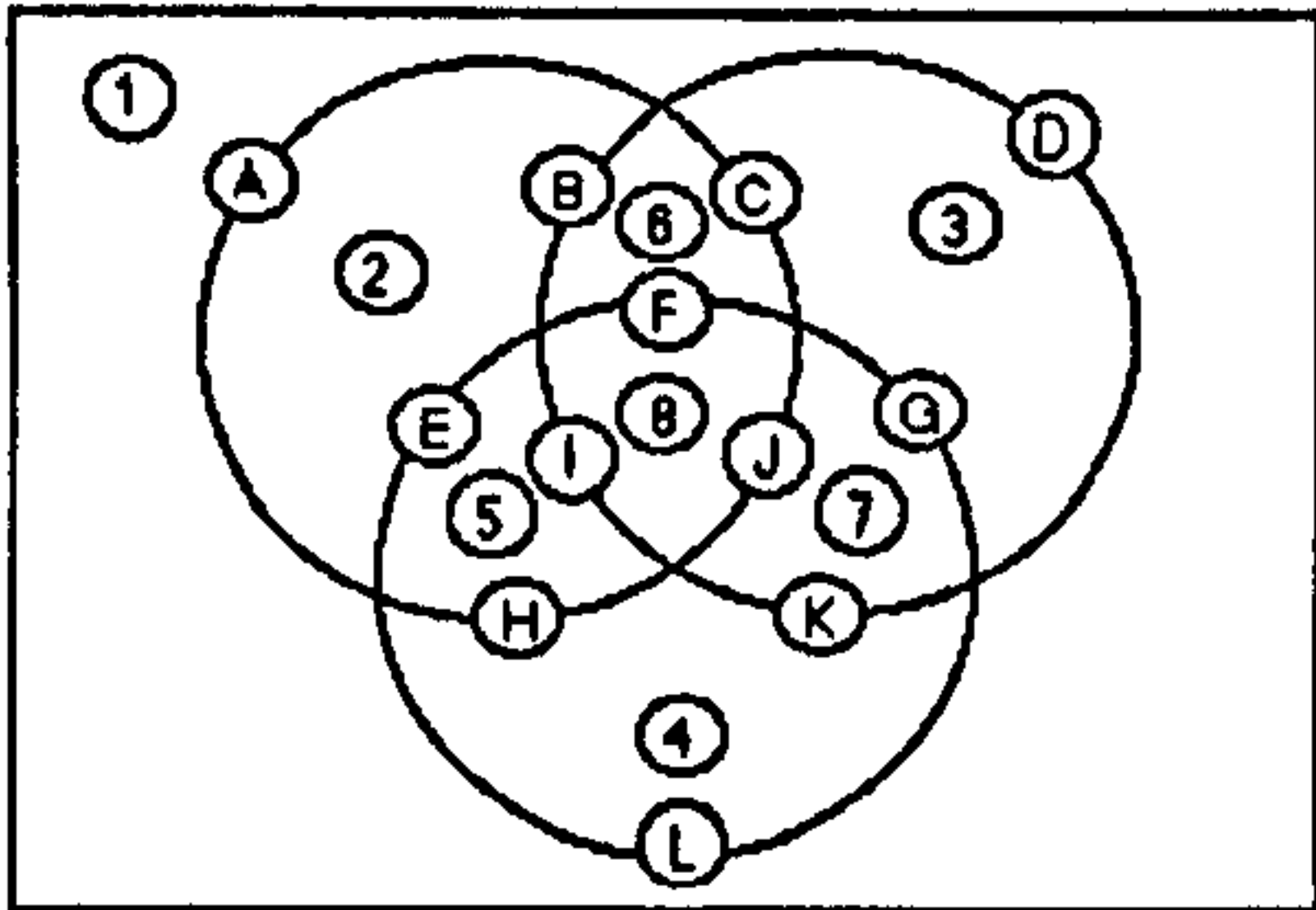
A number of distinct categories of errors and misconception occurred and can be traced through the diagrams that learners produced. This data also points to the differences between the computer based systems and the paper based versions. Many of the mistakes made with Euler are not possible with the current software implementation. The EULER system constrains the user from making registration diagrams that are inconsistent with the premises, although it allows flexibility with respect to intersections that are possible. This restriction



may be useful as a scaffolding device in providing apprenticeship much like the BELVEDERE system (Levonen & Lesgold, 1993), GIL (Reiser *et al.*, 1992) and Burton and Brown's work in subtraction (1978). On the other hand, these constraints may artificially simplify the solving process so that learners are unable to perform without the tool. Further work will address this issue and is mentioned in the final chapter.

6.3.3 Bug types in Venn.

In learner's use of the VENN system, seven categories of errors were found. Incidences of each error were grouped and the frequencies calculated. The seven categories of bug types are: (a) the third term error or TT, (b) the star movement error or \*M, (c) the empty set error or ESE, (d) the premise term order error or TOE, (e) the pathological diagram or Path, (f) the spurious representational conventions or SPC and (g) representing the conclusion as a premise or RPC. A description of each of the buggy-rules follows with externalisations which are indicative of each rule. To more easily show the annotations expected from learners operating with each bug the following diagram was created. Regions in the diagram are labeled 1 through 8 and arcs A through L. The diagram and the labels are used to indicate clearly the areas on the diagram where learners would be expected to make an annotation if they were operating with one of the bugs.



These individual bugs will be looked at in detail below. However they are demonstrated by subjects as relatively unstable and composite behaviours. The process of determining these errors from the annotations made by learners was an exploratory and iterative activity. The data was scanned several times over until certain patterns began to emerge.

*The third term error (TT).*

This bug can be described as a failure to properly annotate a term circle which refers to the term not mentioned in that premise. This error results in missing shading regions and misplaced existential annotations. There are a limited number of possible instances of this error, subcategorized for each of the four premise types. In a universal affirmative, or A mood premise, the diagram should show shading in both regions which complement the

second term circle, including the region which is part of the third term circle. The order of the terms in the premise and the premise order is important, yielding four possible diagrams. In positive existentially quantified, or I mood premises, when the star should be placed on the arc forming part of the ambiguous term circle, the error will force placement of the star in either of the ambiguous regions adjacent to the arc. The adjacent arcs will depend on the premise order, that is if the premise is an A-B premise or a B-C premise. In the case of a negative universal premise, or an E mood premise, the shaded region should cover the intersection between the two terms including the part of the third term circle which intersects with that region.

Table 18: Expected third term error annotations.

		Term order			
		A-B	B-A	B-C	C-B
All A are B	A	S2S5^	S3S7^	S3S6^	S4S5^
No A are B	E	S6S8^	S6S8^	S7S8^	S7S8^
Some A are B	I	*6, *3	*6, *3	*7, *8	*7, *8
Some A are not B	O	*2, *5	*3, *7	*6, *3	*5, *4

*Note.* S = Shading, \* = \* Annotation, Numbers 1-8 identify regions, Letters A-L identify arcs. The symbol ^ indicates that the annotation must be absent from the diagram.

Lastly if a premise is negative existentially quantified, or an O mood premise, then the star notation would be placed on the arc which forms a part of the third term circle within the circle of the first term in the premise. There are more possible regions in this scenario as, unlike I mood premises where A-B term order is equivalent to B-A term order, the order of the terms in O-mood premises has a bearing and yields eight possible diagrams for single premises.

Table 18, should be read as showing all the possible diagram annotations which might be produced, if the subject was displaying this error. The first entry shows that for a problem in the A mood with A-B term order (All As are Bs), the error would predict shading in area 2 and missing shading from area 5 (area numbers refer to the diagram at the beginning of this section).

*Star movement error (\*M).*

This bug applies to a number of situations where the \* annotation is moved from its original position of ambiguity with respect to another term, into another region when one of the regions of ambiguity is precluded by the shading annotation of the second premise. The bug shows a misunderstanding about the nature of one constraining principle in the system, that an existential annotation cannot be ambiguous between two regions where one region is shaded. The error only applies to problems where there is an existential annotation in the first premise and a universal shading annotation in the second premise. There are then only 9 instances from the 64 possibilities where this bug can occur.



Table 19: Expected star movement error annotations.

2nd premise	1st premise			
	Some As are Bs	Some Bs are As	Some As are not Bs	Some Bs are not As
No Bs are Cs	*FS8S7	*FS8S7	-	-
No Cs are Bs	*FS8S7	*FS8S7	-	*GS7S8
All Bs are Cs	-	-	-	*GS6S3
All Cs are Bs	*FS6S3	*FS6S3	*ES5S4	-

Note: \* = \* Annotation, Letters A-L refer to arcs, S = Shading, Numbers 1-8 refer to regions.

Table 19 shows the expected annotations for subjects with this form of error. The first entry shows that when the first premise is Some As are Bs, and the second premise is No Bs are Cs, the expected annotation would be for a \* on arc F, shading in region 8 and shading in region 7. Numbering and lettering of the regions and arcs correspond to the diagram at the beginning of the section.

Empty set error (ESE).

This bug occurs after both premises have been annotated in the diagram. At that point for those problems described in Table 20, there is only one region of a term circle which is not shaded. In this instance the system allows the addition of a further existential annotation to that region. This convention is allowed as we assume that there can be no empty set in the syllogism. When this occurs we can derive a conclusion based on the position of the \* in the diagram.

Table 20: Expected empty set error annotations.

2nd premise	1st premise			
	All As are Bs	All Bs are As	No As are Bs	No Bs are As
All Bs are Cs	*8^	*8^	*6^	*6^
All Cs are Bs	-	*6^	*7^	*7^
No Bs are Cs	*6^	*6^	*3^	*3^
No Cs are Bs	*6^	*6^	*3^	*3^

Note: \* = \* Annotation, Numbers 1-8 refer to regions. The symbol ^ indicates that the annotation must be absent from the diagram to infer the bug.

Table 20 shows all the possible instances of this error. In the first entry, where All As are Bs and All Bs are Cs, the expected annotation from a learner with this bug would be for the \* to be missing from the central region, region 8. Region 8 is the central region according to the diagram at the beginning of this section. In all cases the error will be shown by the absence of a \* in some region.

*Premise term order error (TOE).*

This error occurs partly due to misreading the premises before entering them into diagram. The bug occurs most frequently when the syllogism is in any of the second third and fourth forms or in other words when the premise term order is not A-B-B-C. The counting of this error may be expected to be similar for each system as it is less related to the system itself than to the interpretation of the premises. It may be however that different systems are more or less difficult to translate from the textual versions so still we may expect different degrees of this error from each representational system.

Table 21: Expected premise order error annotations.

Moods	1st and 2nd premises			
	A - B	B - A	B - C	C - B
A	S3S7	S2S5	S4S5	S6S3
I	-	-	-	-
E	-	-	-	-
O	*G	*E	*H	*C

*Note:* \* = \* Annotation, Numbers 1-8 refer to regions, S = Shading in region.

Table 21 shows the annotations which would be expected from this error. The order of the premises is not important in this case and so is the same for both first and second. The mood and the term order determine the 16 possible situations in which this might happen. However in the I and E moods, the annotations are the same if the error is made as they would be if the error was not made. There is no way to tell that this error is being systematically held, from premises in the I and E form.

*Representing the conclusion as a premise (RPC).*

This bug is partly due to the way the problem is presented to the subject. It also reveals an underlying deeper problem about understanding the nature of the systems and the syllogism. The pre and posttests, both present the first and second premises and a conclusion which may or may not be valid. The task for the learner is to solve the problem and evaluate their conclusion against the one offered in the test. Learners who exhibit the RPC type behaviour represent the premises, sometimes accurately on the diagram. They will then represent the offered conclusion as if it were a third premise. Those learners who annotate a diagram with the conclusion statement are seeing the diagram as a set of slots which can be annotated but fail to see that through the annotation any new information will emerge. This behaviour would indicate the more global misconception that the system is not dynamic. All the systems tested here are useful because (through the graphical constraints) they are able to take information (premises) as input, manipulate that input and produce the conclusion as a result. The production of conclusions is complicated by the fact that rules for annotating

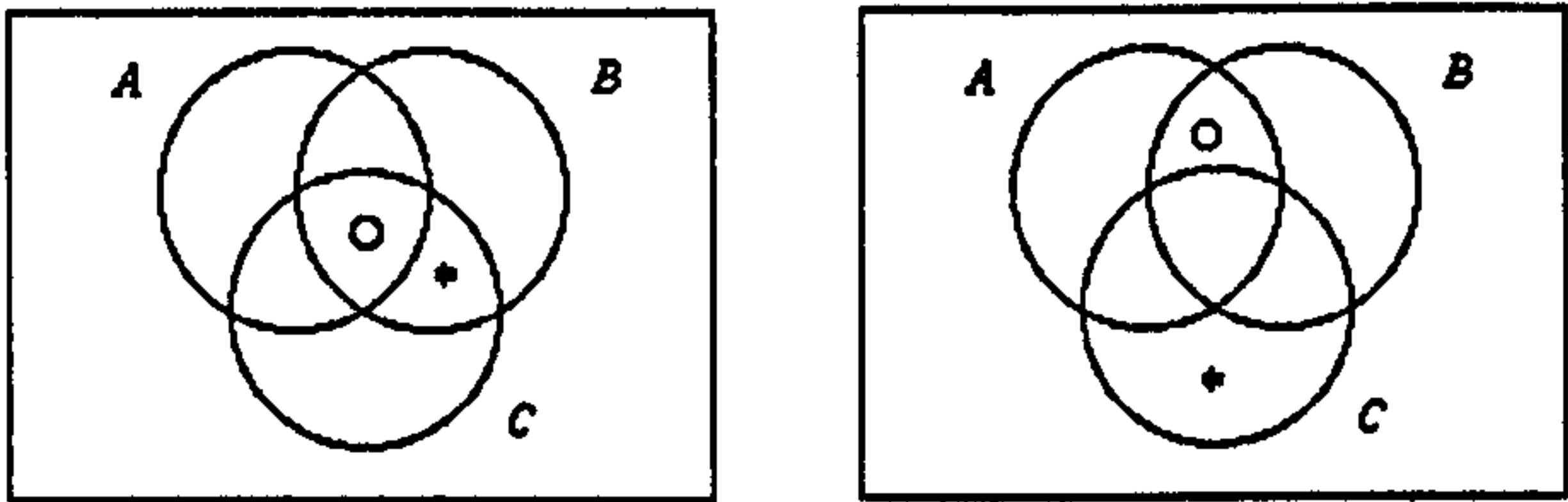


premises are different from those which apply to reading conclusions. The bug might be lessened if the problems were not presented in this way.

The indicators of this error are signs of three sets of annotations on the diagram. That is three stars, three shaded ellipses, three shaded lunules or any combination of these three conventions. In the normal solution process only two annotations are required and the third phase requires interpretation of the diagram which results from annotating the first two premises. A table of possible diagrams and the annotation expected for this bug would be dependent on the conclusion which was offered to the learner. This table is in principle, quite simple to construct, but has not been developed in this project.

*Spurious conventions (SPC).*

In some cases learners create conventions which were consistently used through the range of problems, but were spurious and unable to help them come to a solution to the problem. There are two kinds of invented conventions, ones that work and those that couldn't work. Of those that couldn't work, we have also to make the inference that a higher level misconception is present. This is, like in the RPC errors, the misconception that the systems are not generative or able to produce new information in any way.



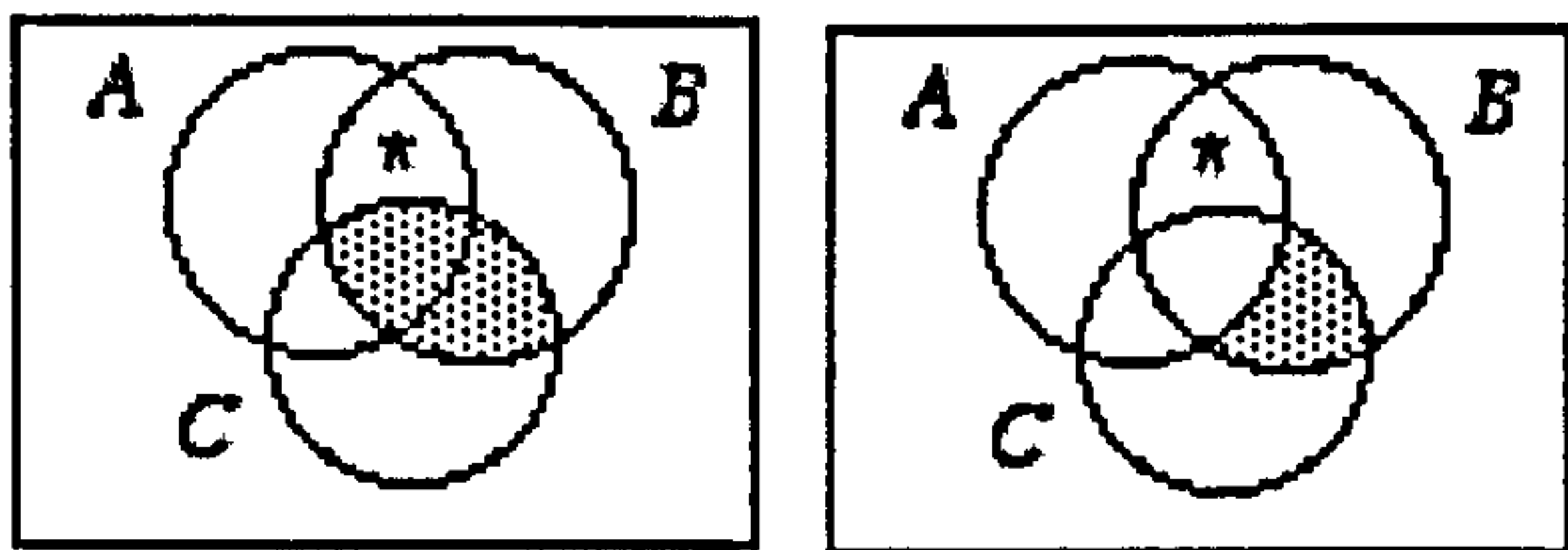
In these two diagrams the subject has created a novel convention which has no relation to the system and will not help the subject to solve the syllogism. The 'O' for this subject was consistently used to show universally quantified premises. In the right-hand diagram the premise All As are Bs is represented by placing the 'O' in the intersection of the A and C term circles. The learner has consistently used an invalid and self generated convention which will not help them solve the problem.

*6.3.4 Diagnosis in Venn*

In practice of course, bugs are not exhibited with stability by the same learner or easily disambiguated from each other, making diagnosis and appropriate remediation difficult. Instability is shown by a lack of a complete correlation between problem type and the annotation which the learner makes. It is therefore difficult in some cases to make a corresponding diagnosis of a bug type in a certain individual when they do not show the same external signs for similar problems. A solution to diagnosis from unstable symptoms, might be to base interventions on probabilities rather than fact. If there are two competing causes, for example, an intervention might be based on the most probable cause. Information

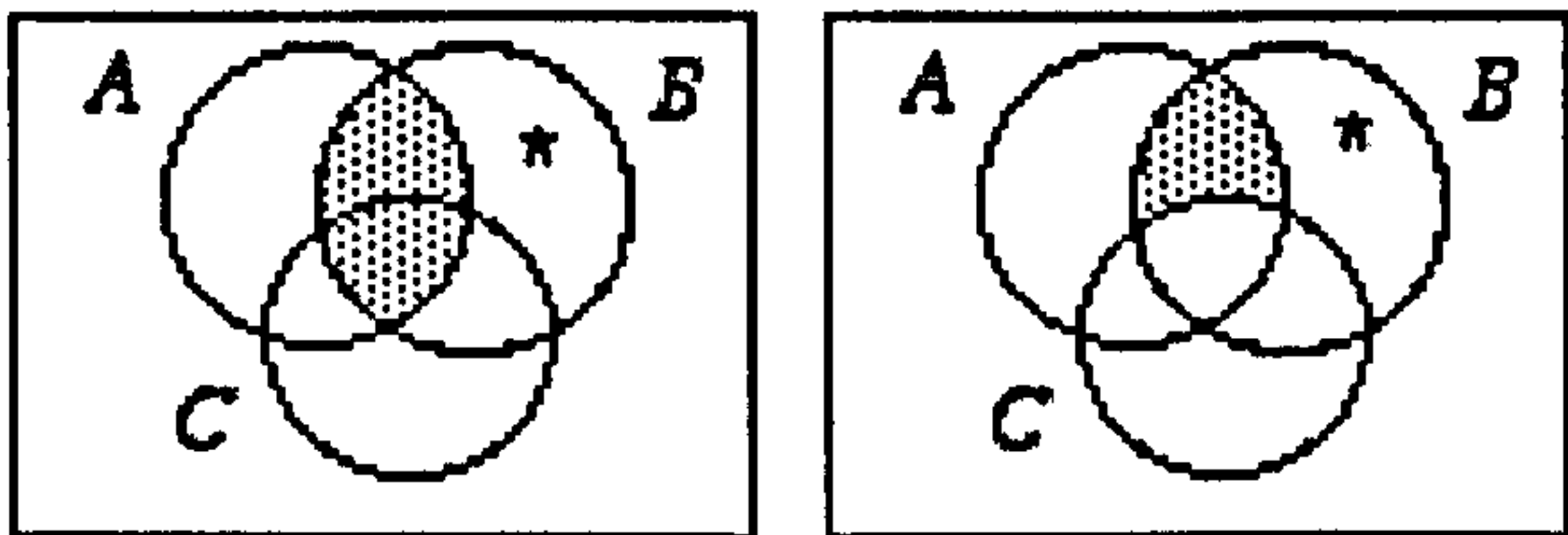
which could contribute to the probability could range across many variables, including features of the population sample, frequency of the symptom, the already known abilities of the individual subject and others. Another problem in diagnosis is composite behaviour. A diagram presented by a learner can indicate the existence of more than one bug. For further disambiguation of particular bugs exhibited in a single learner, a system would need to choose carefully a follow-up question which could yield information to a continually updated learner model. Solution of the diagnostic problem is not the topic of this thesis, however the information provided from this analysis, combined with other data from the literature will help to build a tutoring system for teaching syllogistic reasoning. This is discussed in the final chapter, on further work.

In the following diagnosis, annotations are shown in parenthesis. These codes are taken from the tables of expected annotations for particular errors, above. In each of the diagrams that follow, the correct version appears on the left and the subjects' incorrect diagram on the right. This first diagram, D1, is a simple diagnostic problem where the learner shows a third term error in the first and second premise. The problem was, "Some As are Bs, No Bs are Cs." The learner has located the star in region 6 from the first premise indicated as \*6 in Table 20 and failed to shade the central region indicated as, S7S8^ in Table 18.



D1

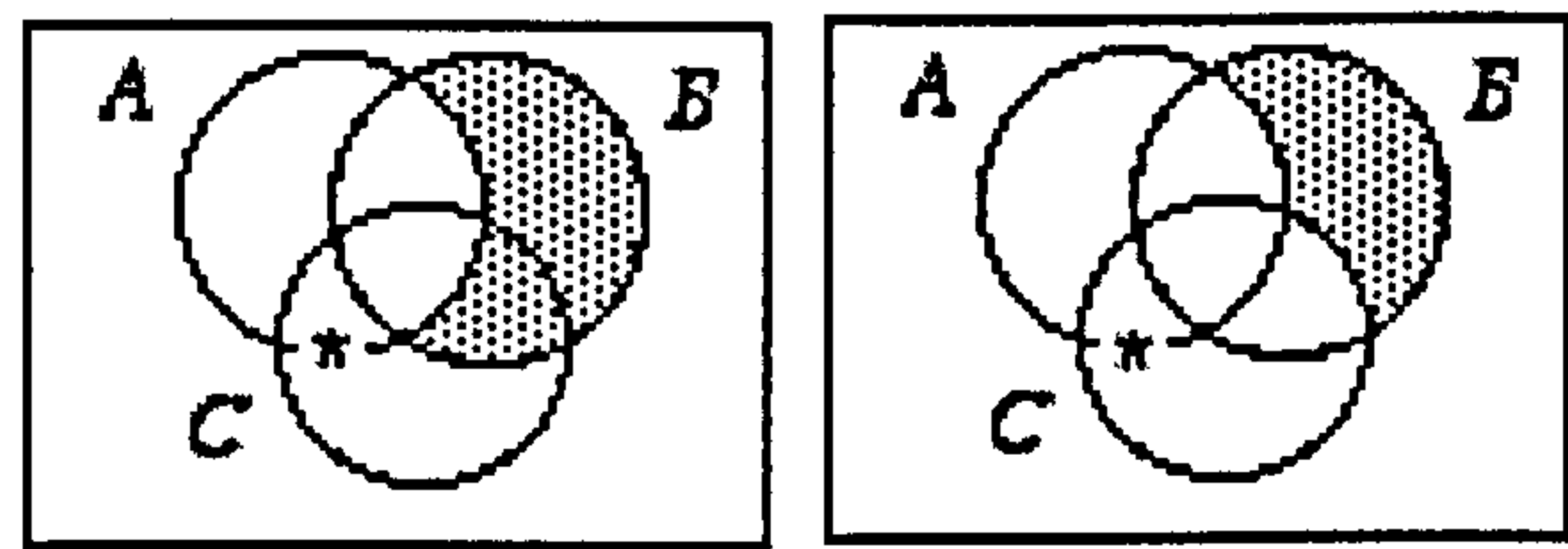
In D2 there are two errors. The TT type error is accompanied by a TOE type error. The syllogism was, No As are Bs, Some Bs are not Cs and the conclusion, Some Cs are not As. This learner failed to shade the central region (8) to show fully the preclusion of individuals which are A and B. The star has been correctly placed in region 3, having been displaced from arc C. There is no valid conclusion to this problem, except the U shaped conclusion that Some not As are not Cs. This type of conclusion has not been discussed in this project.



D2

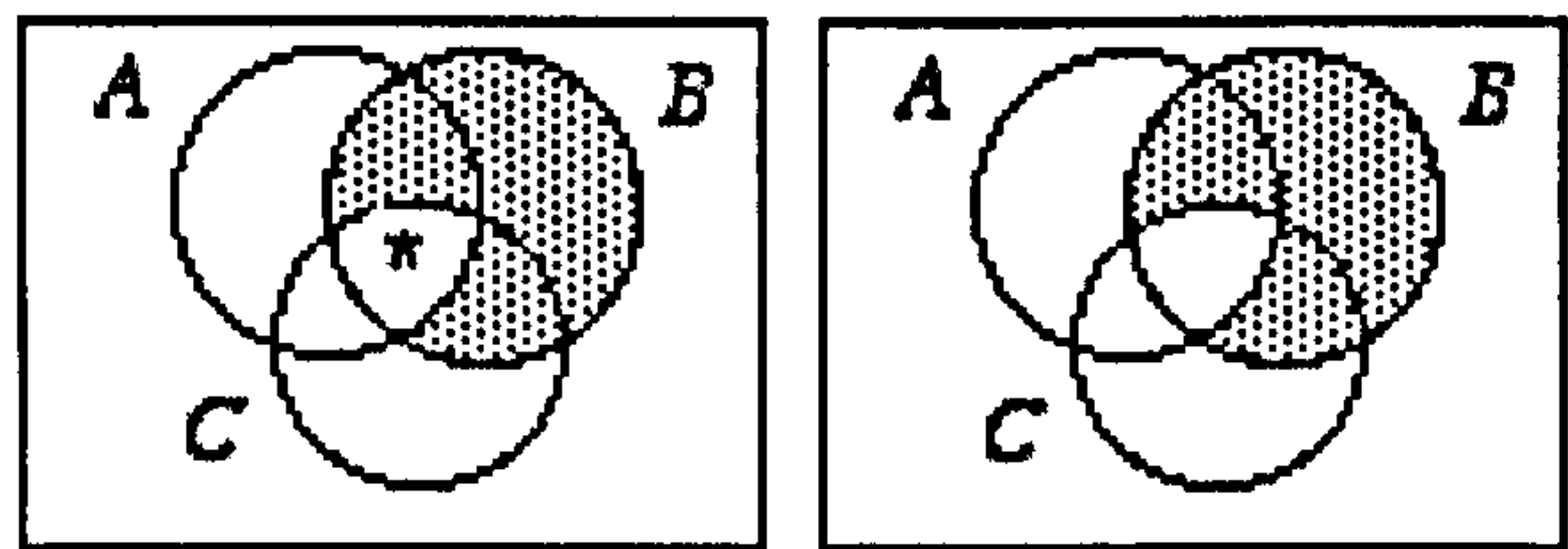


In D3, a TT type error is shown where the problem was, All Bs are As, Some Cs are not Bs. The shading pattern S2S5^ indicates a TT type error for this problem as shown in Table 18. The location of the star is correct for this problem.



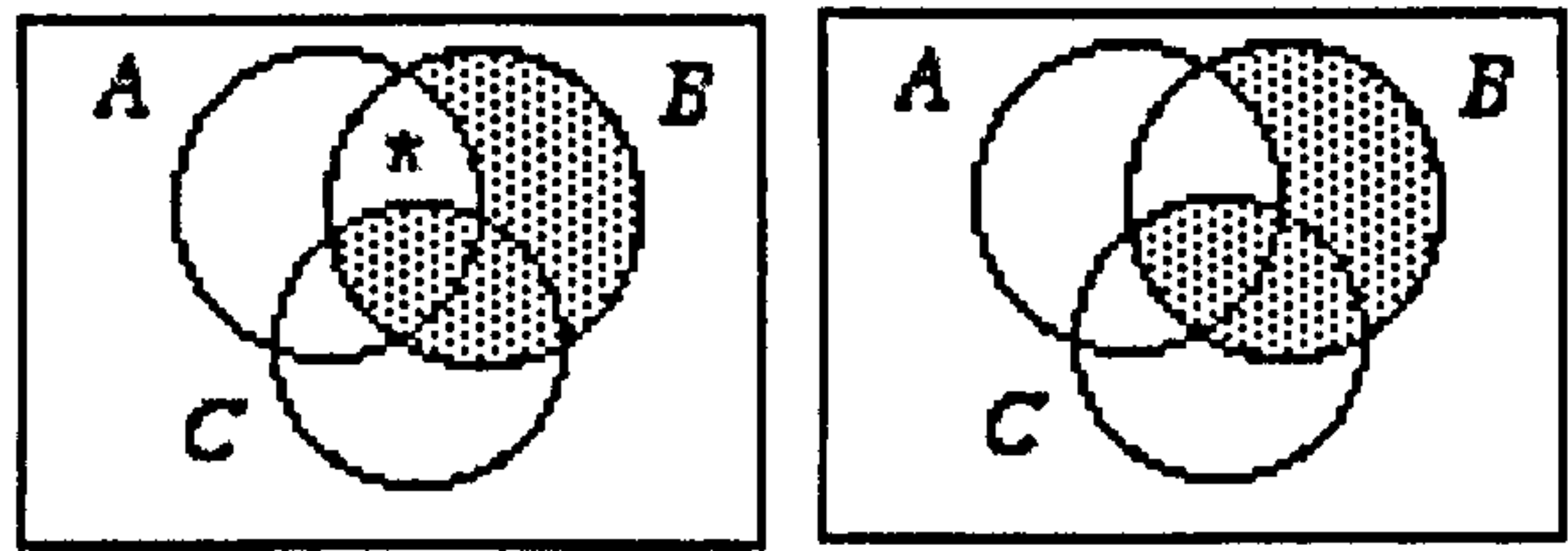
D3

In D4 an example of an empty set error is shown. There should be a star in region 8 for this problem, All Bs are As and All Bs are Cs, so that the conclusion can be read, Some As are Cs. The star should be in the central region, because there is only one region left of B, which is not shaded. The traditional syllogism claims that there can be no terms which have no existing members. This is sometimes called the ‘no empty set assumption.’ Because of this assumption, an existential annotation should be made to the central region, region (8). Once the central region has a \*, the relation between the C and A term shows they have individuals in common. This observation allows the inference, Some As are Cs, or even, Some Cs are As.



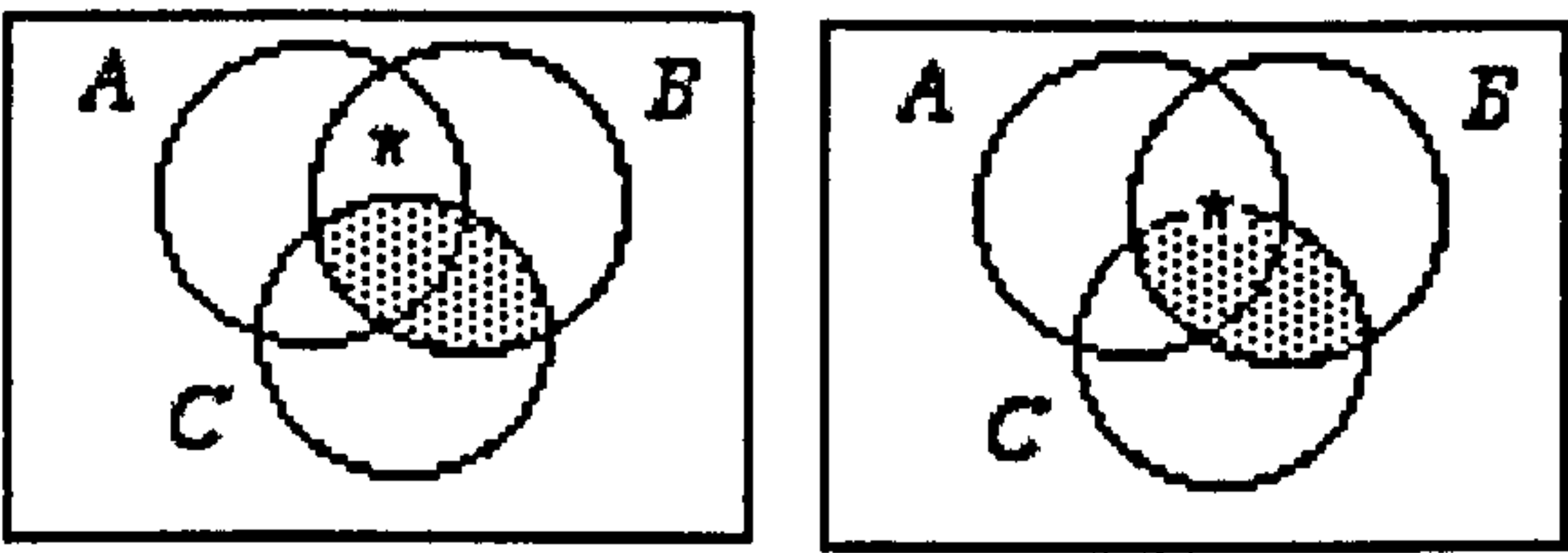
D4

Diagram D5 shows the syllogism, All Bs are As, No Bs are Cs. The diagram should have a star in region 6, so the conclusion can be drawn that Some As are not Cs. A TT type error is diagnosed.



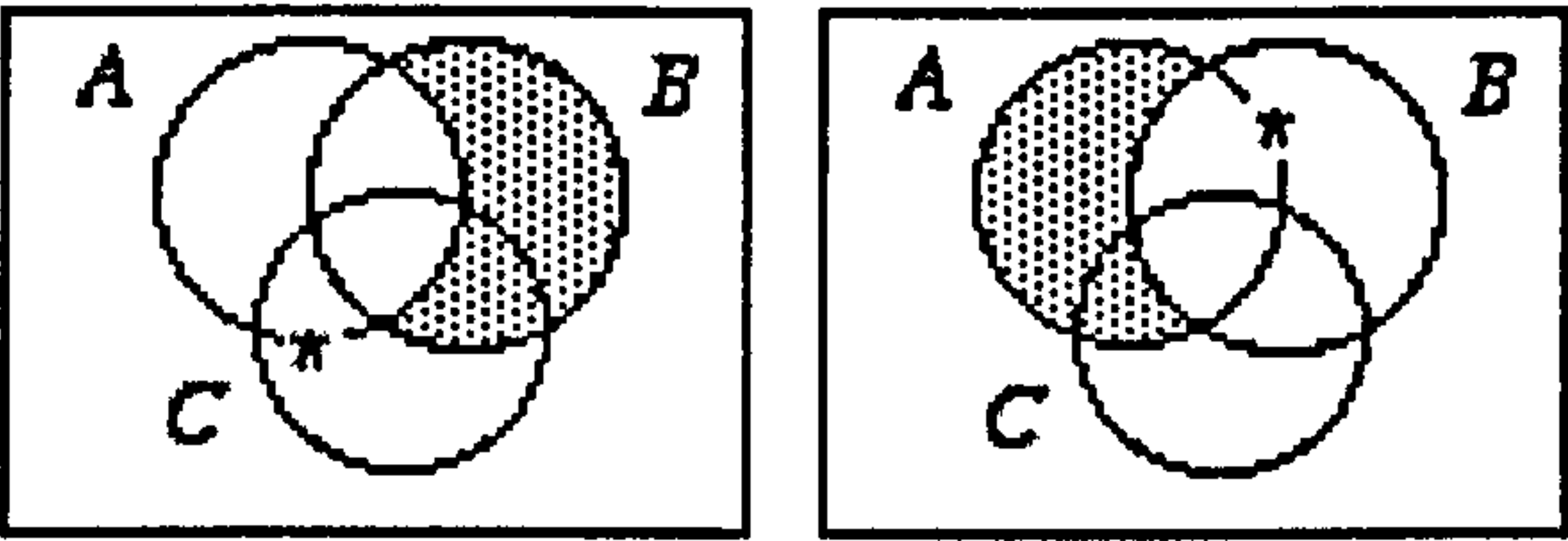
D5

In D6, the first premise Some As are Bs is correctly annotated however, when the second premise is added, No Bs are Cs, then the star should have been moved into region 6 directly above thus indicating a \*M type error as shown in Table 19.



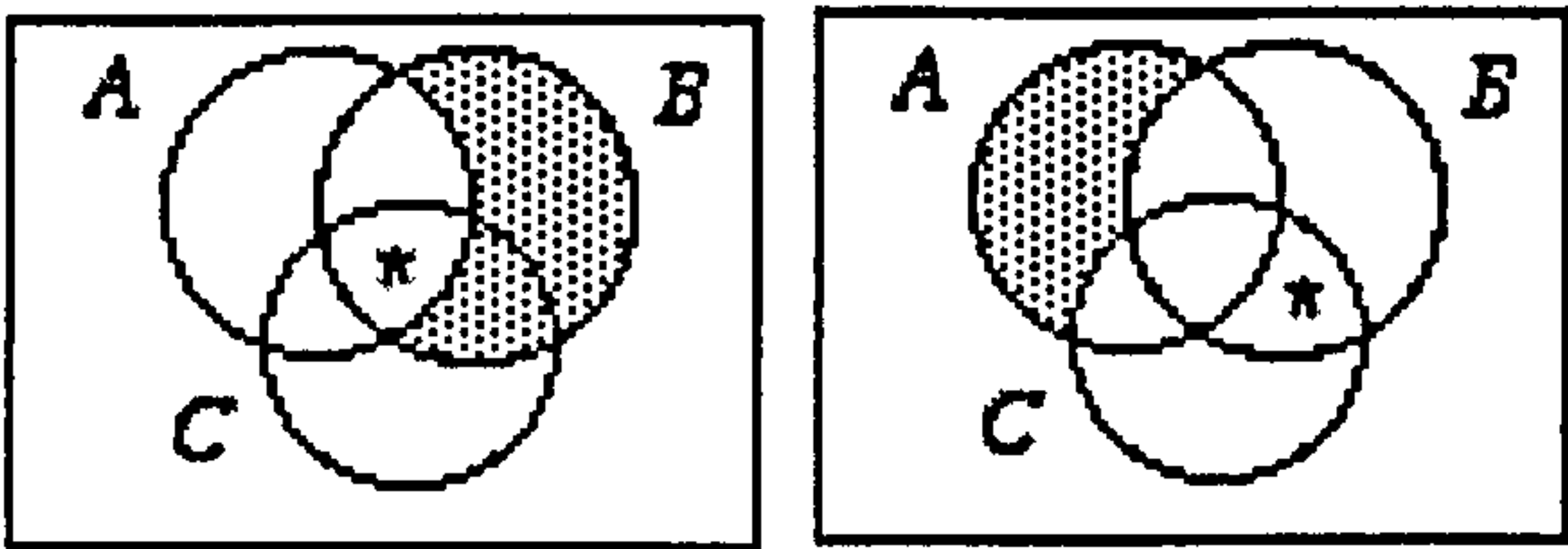
D6

In D7, the first premise, All Bs are As has been annotated, All As are Bs yielding S2S5 as shown in Table 21 and implying a TOE type error. In the second premise Some Cs are not Bs the order has again been reversed. This suggests the learner is systematically reading A-B, B-C term order into each problem. The subject is showing double evidence of TOE type errors.



D7

In D8 a compound of errors exists comprising systematic misreading of term order, a TOE type error and third term, or TT type error. The problem was, All Bs are As and Some Bs are Cs. The learner has shaded region 2 indicating that they read the first premise as All As are Bs and then missed the shading from region 5 indicating a third term error. This later diagnosis is reinforced by the placement of the star from the second premise. The second premise would have been annotated similarly regardless of reading since it is in the I mood so the misreading diagnosis is no further reinforced, however the placement of the star in region 7 indicates a TT type error. The combination of TOE and TT type errors gives in this case S2S5^ plus S2S5^ leaving S2 as shown in the diagram.

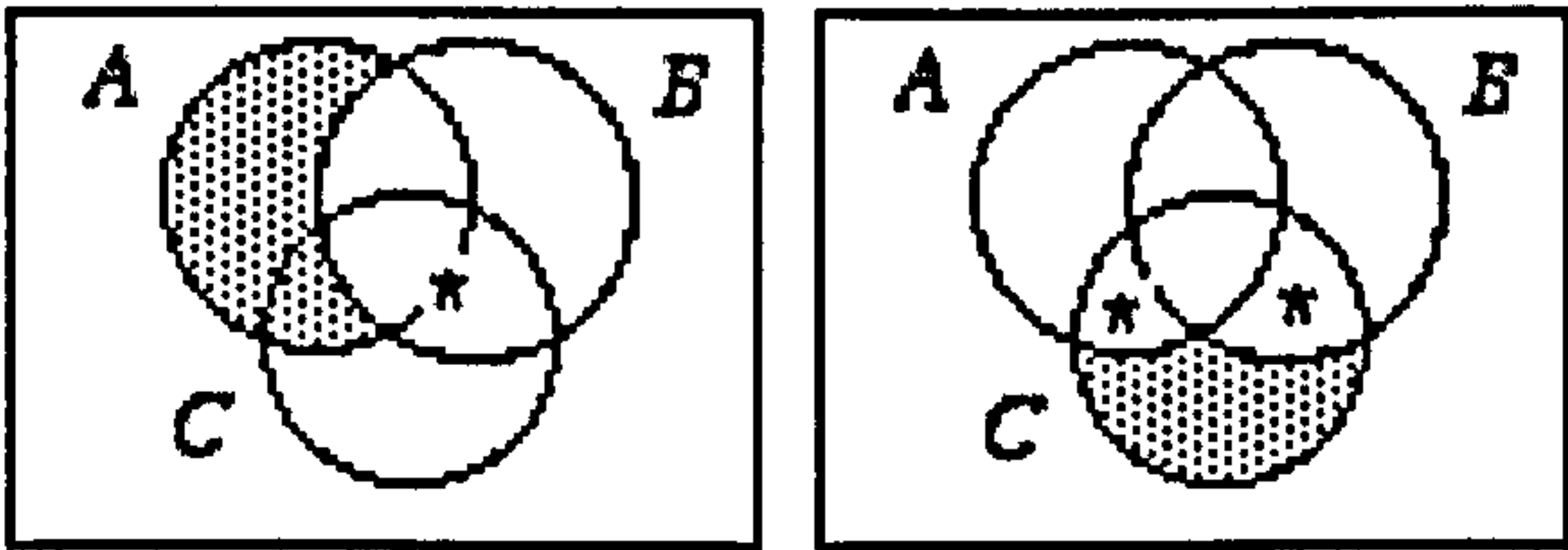


D8

Unfortunately some diagrams remain impossible to interpret even according to the combinations of buggy rules and errors shown. Diagram D9 for example shows what has been called a pathological diagram or Path type error having no relation to any of the bugs

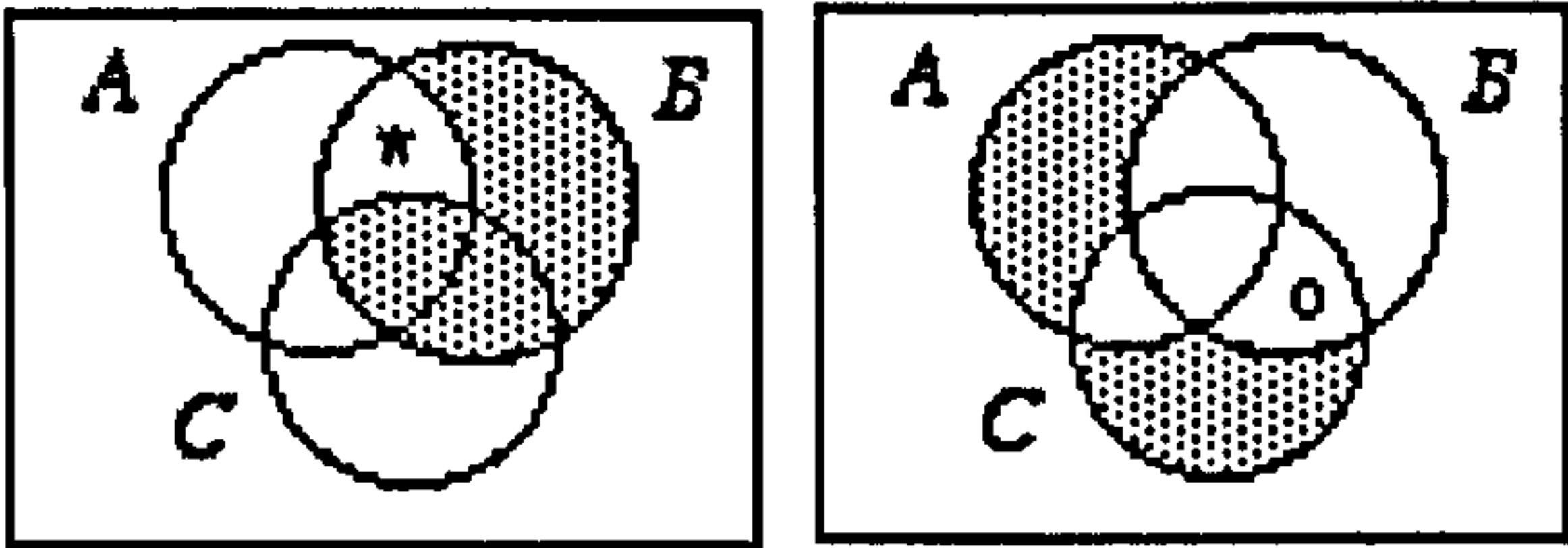


or the correct annotations. There are signs of third term or TT type error since the shape of the shading is reminiscent but not enough to establish this for sure. The premises were All As are Bs and Some Bs are Cs. The diagram also includes three lexical items, two stars and 1 shading. This may indicate that the subject is trying to annotate the conclusion instead of reading it from the diagram, an RPC type error.



D9

Finally, D10 shows three annotations have been used incorporating the conclusion an RPC type error, misreading the term order, a TOE type error and inventing a convention or SPC type error. The problem was All Bs are As, No Bs are Cs and the offered conclusion was All Cs are As. The subject has used shading to correctly show the first premise but with third term error has missed region 5 leaving S2S5^ . The subject invented the annotation of the small circle to show nonexistence between B and C and then shaded the conclusion, All Cs are As again with the disbenefit of a third term error so not shading region 7.



D10

These diagnosis were selected from the data set as progressively more complex examples to illustrate the analysis. The same technique was used on all of the diagrams produced by the learners for the VENN system and for the EULER system presented in the below. Each of the 165 diagrams produced by subjects were collated, frequencies were noted for each of the discipline groups.

6.3.5 Bug distribution in Venn.

There are seven major bug categories produced by learners. These are bugs which can be assessed from the diagrams themselves. There are almost certainly others which were not found from this analysis and an interview technique might determine those. Exactly 165 VENN diagrams were provided in posttest by subjects and of those, 105 were provided by mathematics and 60 by humanities students. Table 22 shows the incidence of each bug and

error type as a raw score and as a fraction of diagrams generated by the particular discipline condition. Since each bug has been identified to a particular graphical construction and each graphical construction to an individual in a group we can map the bug type incidence to the learner group as follows.

Table 22: Venn error spread between learner groups.

	Bug types						
	SPC	RPC	TT	TOE	Path	*M	ESE
M (d=105)	0 (.00)	0 (.00)	9 (.09)	6 (.06)	2 (.02)	6 (.06)	27 (.26)
H (d = 60)	7 (.12)	14 (.23)	15 (.25)	15 (.25)	17 (.28)	0 (.00)	0 (.00)

*Note:* SPC = Spurious conventions , RPC = Representing the conclusion , TT = Third term error, TOE = Term order error, Path = Uninterpretable diagram, \*M = \* Movement error, ESE = Empty set error. M = number of diagrams produced in mathematics condition, H = number of diagrams produced in humanities condition.

For each group of students a different set of bugs were shown. Bugs have been grouped into those shown only by humanities students as group one. These were the SPC and RPC bugs. Those bugs shown by both mathematics and humanities students group are shown as group two. These are the TT, TOE and Path bugs. Bugs only produced by mathematics students are shown as group three. These bugs are \*M and ES. A  $\chi^2$  test of the categories shows that although most bugs are common to both groups there is a significant difference in those categories between mathematics and humanities students ( $p < .001$ ). We see that humanities students display different behaviours to mathematics students. Humanities students tend to create their own conventions without possibility of creating a successful tool. This behaviour was not exhibited at all by any mathematics students. The behaviour of representing the conclusion as a premise was demonstrated in 23% of all diagrams generated by humanities students. This shows a misconception at the level of treating the diagram not as a mechanism for which there can be a conclusion derived but as a picture into which information can be filled.

Bugs shown by both groups of learners (TT & TOE) were seen more often in the non-mathematically trained learners. The third group of errors, demonstrated by only mathematics students, included moving the star, and the empty set error. These are “advanced errors” that tend to occur at late stages of the algorithm. Learners showing these errors have clearly progressed further than learners showing more naive errors.

The distribution of these bugs is consistent with the general result that mathematics students did much better at learning with the software than the humanities students. These results also show a progressive nature in the differences between bugs. Some of the bugs are elementary, and others are signs that most of the procedure has been apprehended. The \*M



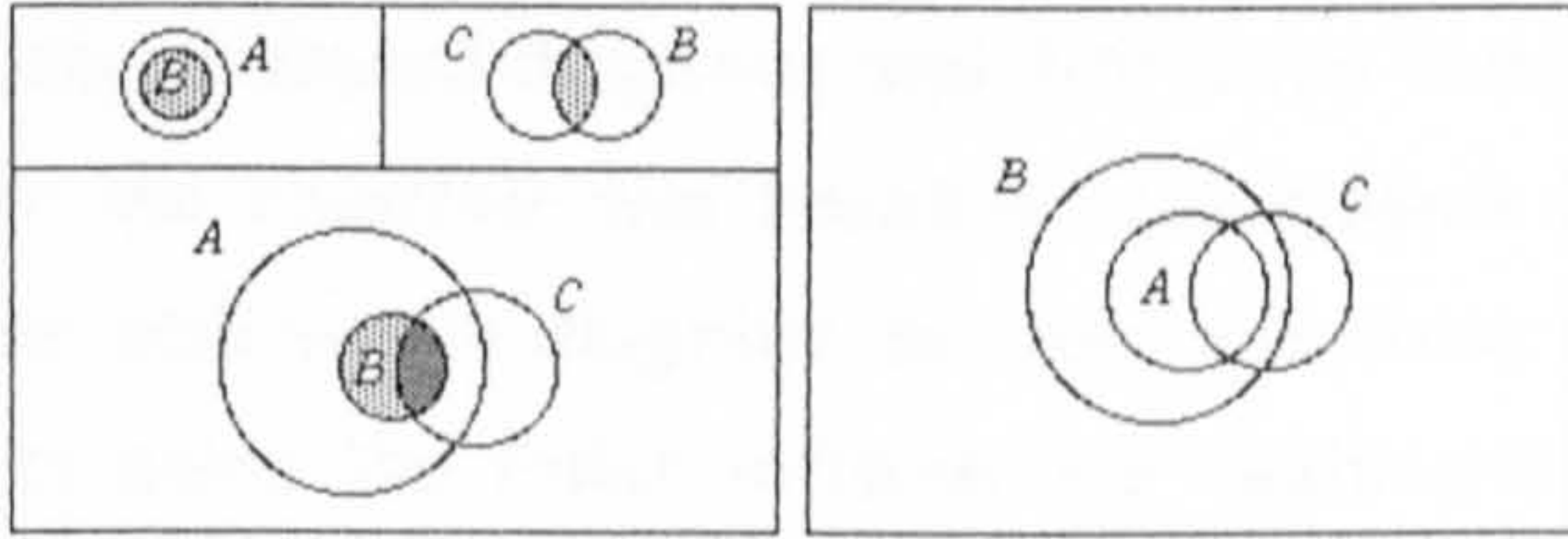
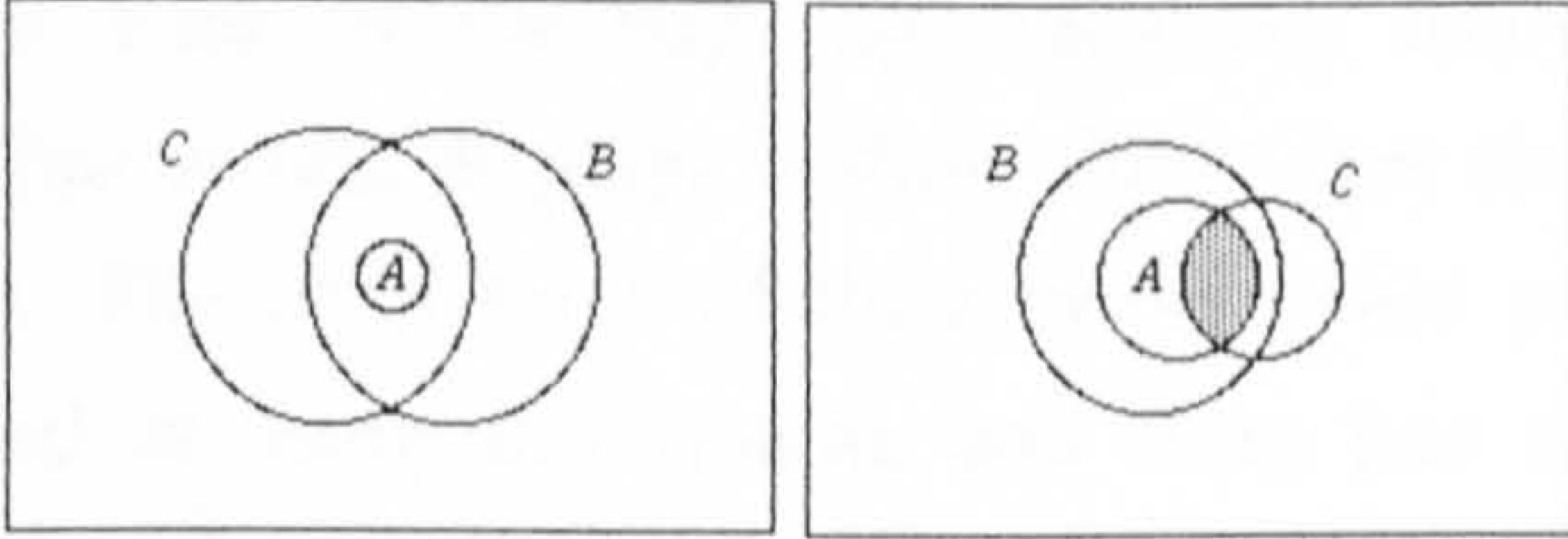
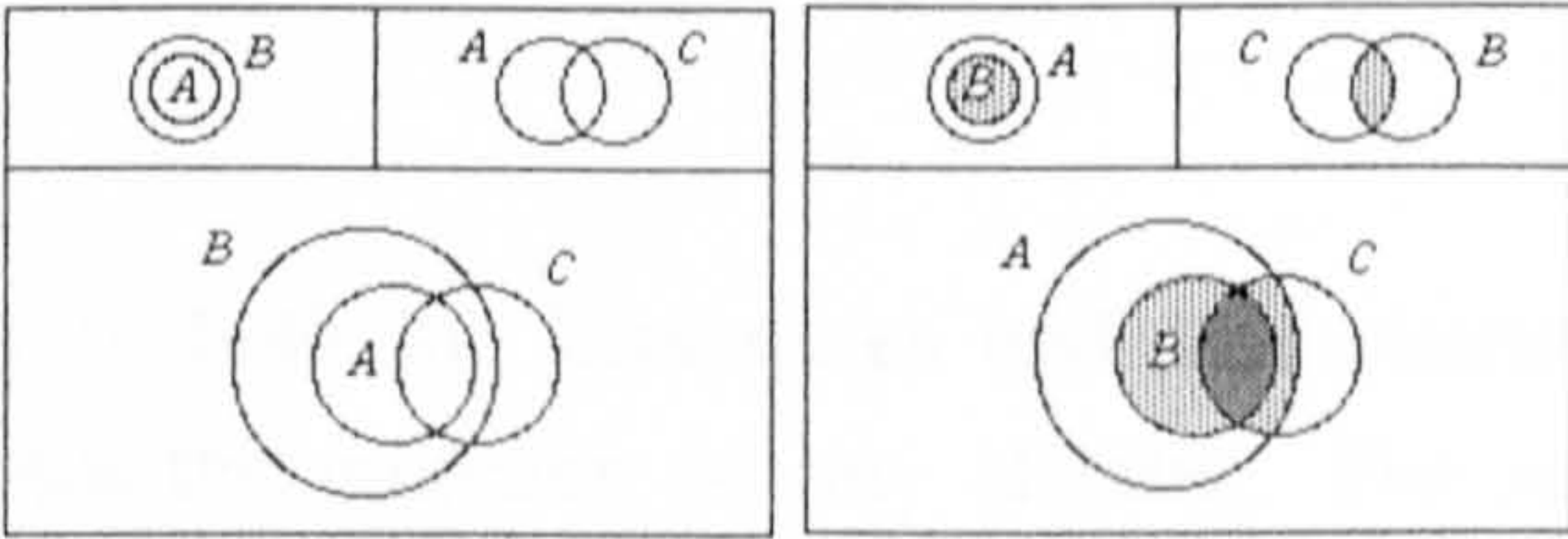
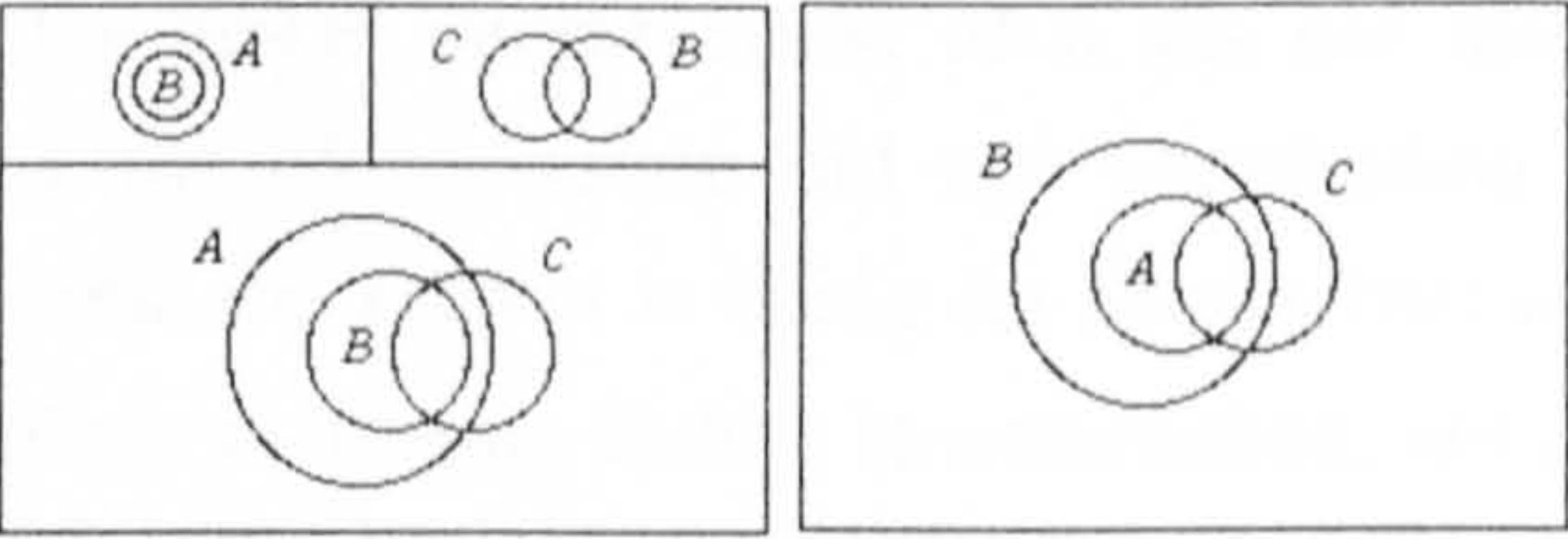
bug, for example, is only important in 9 of the 64 problem types. A student could answer very many of the problems correctly without coming across the need for this convention.

6.3.6 Bug types in Euler.

A similar analysis was conducted on externalizations generated by the 21 subjects who worked with the EULER software. The investigation generated a similar diagnostic kit where the eight major distinguishable bug types can be shown as follows: a) no shading or NS, (b) no premise diagrams or NPD, (c) incomplete labeling or IL, (d) premise term order error or TOE, (e) missing circle intersection or MCI, (f) pathological, uninterpretable diagram or Path, g) novel conventions which may work or NCW and (h) uninterpretable conventions or UC. Table 23 is a page of data generated from seven mathematics students using the EULER system for a single problem. It shows each of the seven subjects who attempted this problem in the mathematics condition. The problem was, All politicians are liars, Some bachelors are politicians. The offered conclusion that All liars are bachelors is an invalid conclusion. The upper left diagram is the correct diagram and annotations are presented for each diagram showing the diagnostic process.



Table 23: Example data for EULER Mathematics.

Correct version.		S <sub>4</sub> No premise diagrams. No shading. Circle configuration is OK given misreading of terms.
S <sub>1</sub> No premise diagrams. Misreading of B-A relation and hence containment relation in diagram.		S <sub>5</sub> No premise diagrams. Terms must have been misread from premises. Shading shows critical region.
S <sub>2</sub> Mistranslated premise terms. No shading. Registration diagram circles are in correct configuration, given mistranslation. No shading in registration diagram.		S <sub>6</sub> Premise diagrams are correct. Registration diagram is only wrong in shading of $(\neg B \cap A) \cap C$ . Critical region is shaded correctly.
S <sub>3</sub> No shading. Premise diagrams are in correct configuration. Registration diagram circles are in correct configuration. No shading in registration diagram.		S <sub>7</sub> No premise diagrams. Misreading of terms from premises. All A are B, instead of All B are A. No shading.

NS. No shading	(S <sub>1</sub> , S <sub>2</sub> , S <sub>3</sub> , S <sub>4</sub> , S <sub>7</sub> )
NPD. No premise diagrams.	(S <sub>1</sub> , S <sub>4</sub> , S <sub>5</sub> , S <sub>7</sub> )
TOE. Premise term order error	(S <sub>1</sub> , S <sub>2</sub> , S <sub>4</sub> , S <sub>5</sub> , S <sub>7</sub> )
SCW. Novel shading convention which may work	(S <sub>5</sub> )

*Note:* S<sub>1-7</sub> = subject (seven in each condition). This page represents the seven respondents' graphical constructions for a single problem. This posttest syllogism was: All buskers are astute people, All buskers are clever, therefore, All astute people are clever. It is isomorphic and iso-believable with, All business people are astute, *etc.*, in the pretest question illustrated in Table 15.

In this example, five subjects did no shading, four subjects drew no premise diagrams, five read the term order from the premises wrongly and one subject seemed to have invented a shading convention which may work. That shading convention can be seen as a subset of the shading which should have been used according to the standard system. The following data and analysis is representative of the diagrams provided by subjects in conditions (b) and



(e), mathematics and humanities students using the EULER system. The study could potentially have generated 215 diagram items from the seven subjects and fifteen items in the posttest. However not all of the subjects produced diagrams and 109 items were collected.

It is not clear whether the shortfall was because respondents felt confident in their abilities to solve the problems without the diagrams or there was some other reason.

Bugs shown by learners using the Euler software are qualitatively different from those produced by VENN learners. Four of the bugs are omissions from the correct complete diagram. Only one of the four errors, displays without doubt that the learner has an invalid model of the solving process. The other three of these errors could just as well be shown by learners who were very good at using the system, and were just too quick to make the annotations on the diagram.

#### *No shading (NS).*

A high frequency bug in both the humanities and mathematics groups appearing in over 60% of all diagrams was the omission of any shading. The relationship between the traditional Euler system and the newer interpretation of it was not made explicit; however, as has been shown in Chapter 3, the older system did not use shading. It is possible that the shading only makes explicit what the subject is doing for themselves in the head (the likeliest interpretation of the older Euler), indicating failure to externalize, not necessarily a bug in the problem solving process. Several later results comment on the shading convention in these systems. In Chapter 7, two versions of the Euler system were used, one with the shading like the system used in this study, and the other with a star to replace the shading. Learners preferred the star type representation. Chapter 7 comments further on learners' poor use of shading, suggesting some reasons for this problem.

#### *No Premise diagrams (NPD).*

A similar failure to externalize the problem solving process is exhibited when subjects fail to produce premise diagrams. This observation occurs with high frequency in both groups. Nearly half of total diagrams in conditions, (b) and (d) (mathematics and humanities students, using Euler), showed no premise diagrams.

#### *Incomplete labeling (IL).*

A low frequency bug was the omission of labeling from the premise or conclusion diagrams. This occurred when the name of the term or label for the term was omitted from the diagram.

#### *Premise term order error (TOE).*

This error is not solely dependent on the diagram making process. It is also related to the subject's ability to interpret the terms from the syllogism. The process of getting the term order correct in the diagram involves, recognizing which is the middle term (essentially the

term that occurs both in the first and second premises), and then finding the end terms (the others). One might expect the error to be as likely in all experimental groups. The error is inferred from premise diagrams which have more than one A or C term. There can only be a shared B term between the two premises. The EULER software will not allow intersections of circles which are inconsistent with the premises, nor will it allow incorrect or nonexistent labeling.

*Missing circle intersection (MCI).*

Since in the Euler system, circles have to be intersected with each other as part of the problem solving process, subjects can fail to find all the possible intersections. This error is not possible to make while using the EULER software. It is therefore not one which the system can correct for the learner. A tabulation of required intersections for each of the possible problems could be prepared and each diagram could be simply checked against the table entry.

*Pathological case (Path).*

Some diagrams defy all interpretation of either the errors in reasoning or errors in producing the external representation. These diagrams occurred fairly infrequently although mostly in the humanities conditions. The definition of a pathological diagram is one which is incorrect but none of the misconception categories explain the annotations. There will always be room for more investigation of these behaviours so long as they still exist.

*Novel conventions which may work (NCW).*

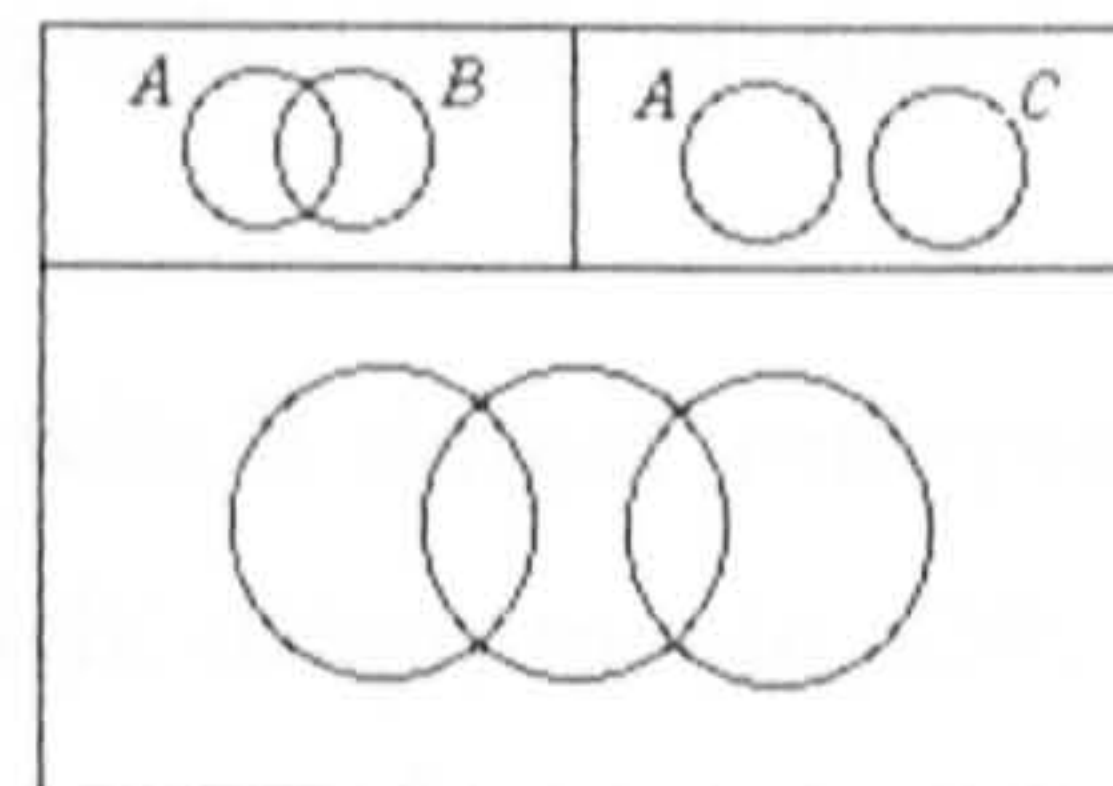
In rare cases, learners might create a convention which did the job of the one which had been taught to them, but was their own creation. This did actually happen for one subject in this study, for one of the conventions in the EULER-mathematics condition. In this case the new convention is a simplification of the convention which was taught. When two shaded areas are intersected, in Euler, the new intersection is often the region which is used to determine what the conclusion will be. This is called the critical region. To distinguish other shaded regions from the critical region, the Euler system usually shades more heavily than its two neighbours. In the software, the intersection is shown as the combination of the two different coloured and different patterned shadings from the neighboring regions. The one isolated learner in this study who created the new convention, made an adaptation based on the software. The learner decided to shade the critical region, and to leave the neighboring regions blank. This convention would work for many of the problems. It is not known whether the learner could have used his convention successfully on all problems since we only have a small selection of problems on which he used the method. On those problems, the learners method was successful.





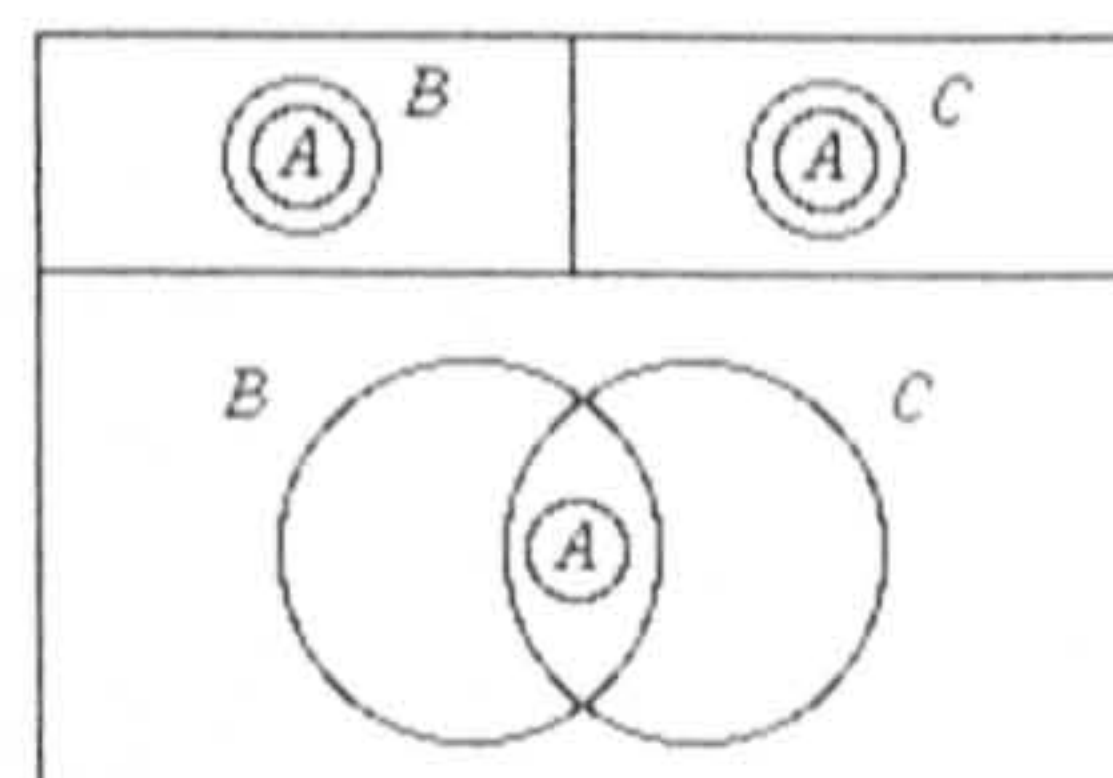


more difficult for the subject to remember the relationships between terms so that the conclusion can be drawn. In both examples there is no shading. The actual premises were for D3, All Politicians are Liars, Some Bachelors are Politicians. There was therefore a misreading of the premises indicating a TOE type error where B should be enclosed by A and not the other way around as shown. For D4, the problem was All buskers are astute and All buskers are clever. Again, there was a misreading of the premises. The B circle should be inside the intersection of A and C.



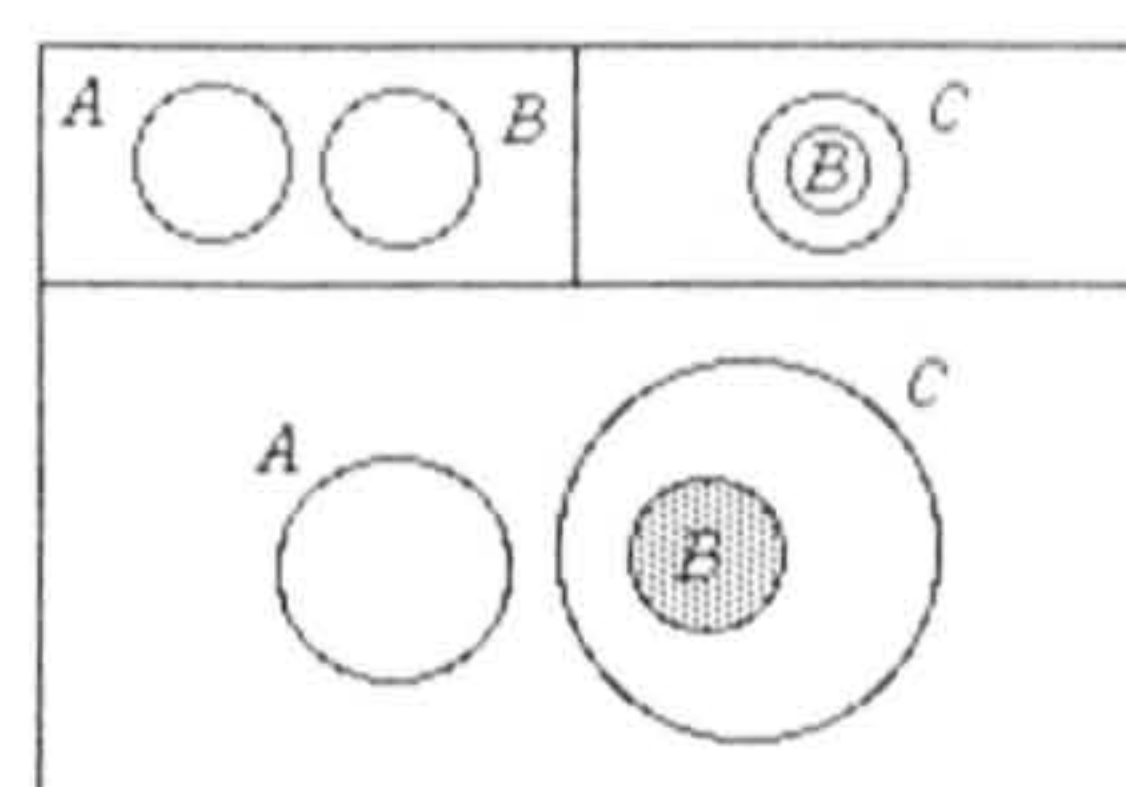
D5

In D5 the subject has not completed the labeling process required to solve the problem. We have called this an IL type error. In the conclusion diagram there is no shading in premises or conclusion showing an NS type error. The A term is represented twice in the premises showing a TOE type error. The premises, Some books are fictional stories, No radio play is a book are represented in the premise diagrams with two A term circles indicating an error in reading the premise term order.

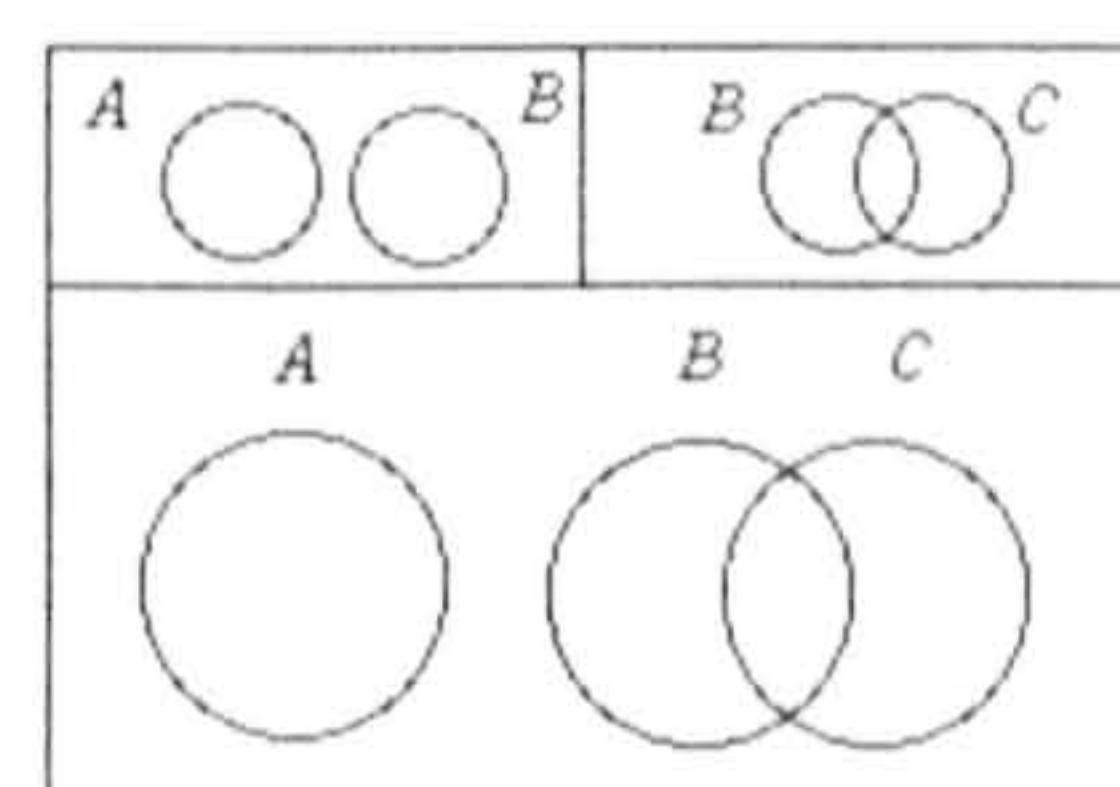


D6

In this problem shown in diagram D6 which is another subject's attempt to solve the problem in D4, All Bs are As, All Bs are Cs, All As are Cs (an invalid conclusion), the subject reads the premises as All As are Bs, All As are Cs reflecting a term order error or TOE type error. The diagram is drawn as if there were two A terms in the problem.



D7



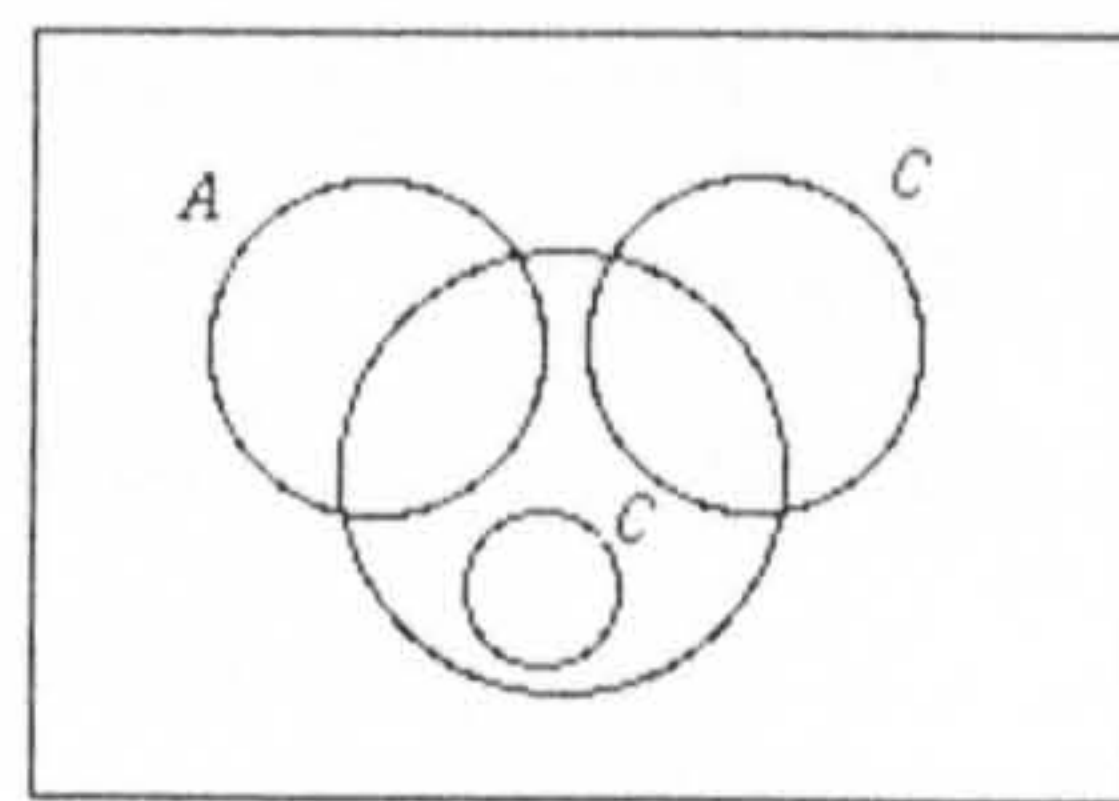
D8

The problem shown in diagram D7 was, No Castles are Mansions and All Mansions are Stately homes. The subject has correctly drawn the premise diagrams although with missing shading in both premises showing an NS type error. He then failed to see the possible

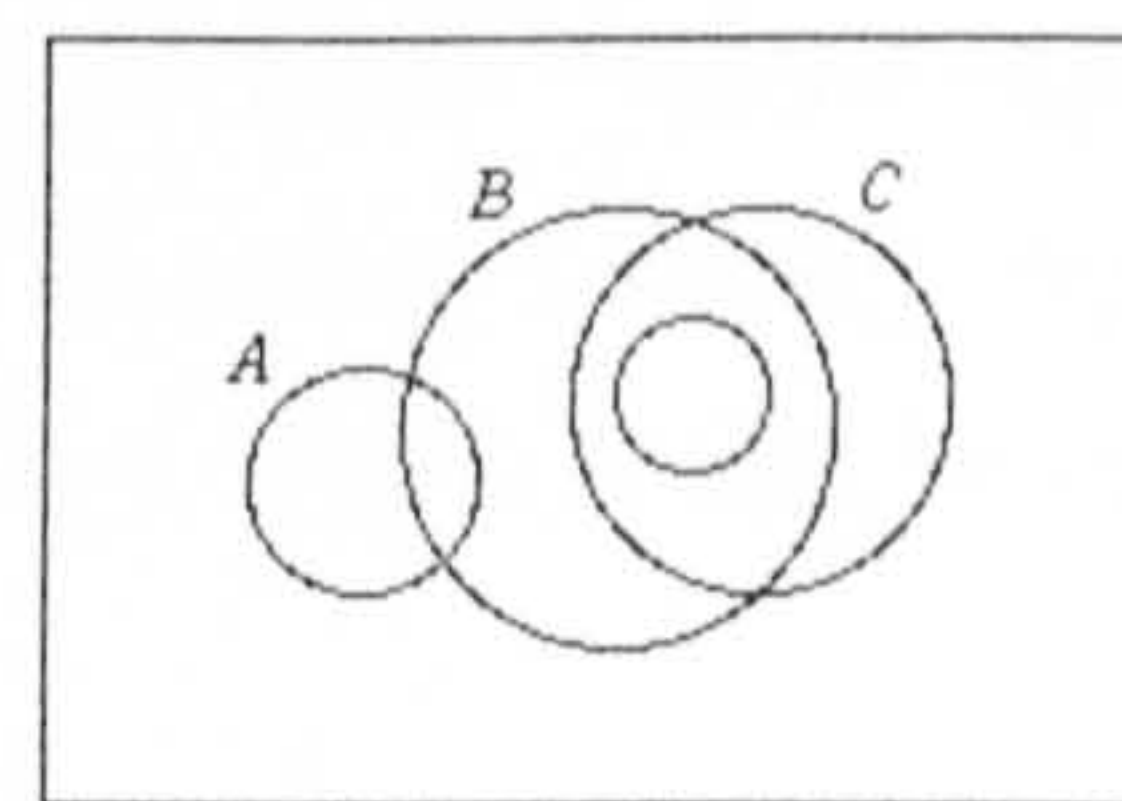


intersections between the A and C term circles showing a MCI type error. Significantly the subject has shaded the critical region in the diagram which leads to the valid conclusion that Some Stately homes are not mansions. This is likely to be an example of the subject taking a short cut in the solution process. In D8 the problem No guitarists are conductors, Some conductors are deaf, is correctly represented in the premise diagrams beside shading, and then the potential intersection between C and A is omitted. This is significant as in this problem compared with one where the first premise terms were reversed there is an additional effort in conceptually shifting the A circle to the other side of the B term circle with which it may not intersect.

Again there are some diagrams which defeat interpretation such as those shown in D9 and D10. There are four circles in each diagram. In D9, two circles have been labeled C. These and similar arrangements have been called pathological or Path type errors. No stable buggy-rules or errors can be induced from them. This behaviour may be an indication of the general misconception about the generative nature of the systems. It may also be that a special buggy rule is going on, which has not been figured out yet. More work with new subjects might help to clarify what is going on for these learners. There is a suggestion in these diagrams that the error has to do with the inability to properly register the B term (or middle term), circles. This would be worth working on in future work and is briefly discussed in Chapter 8.

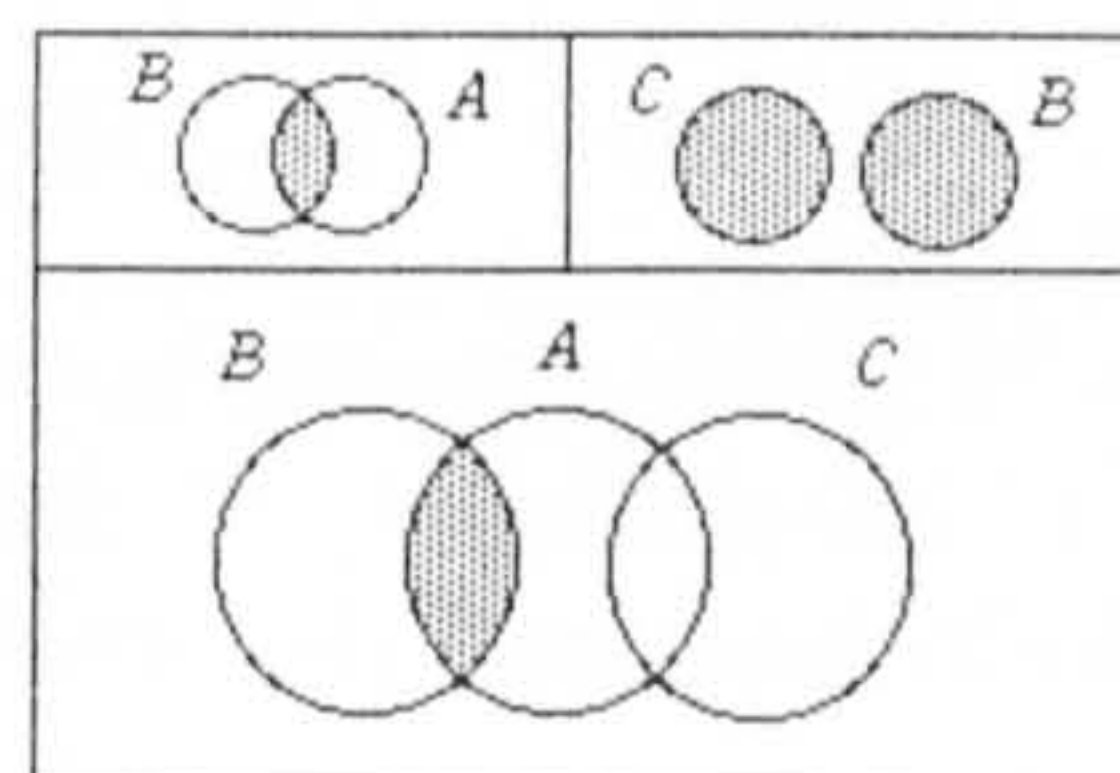


D9

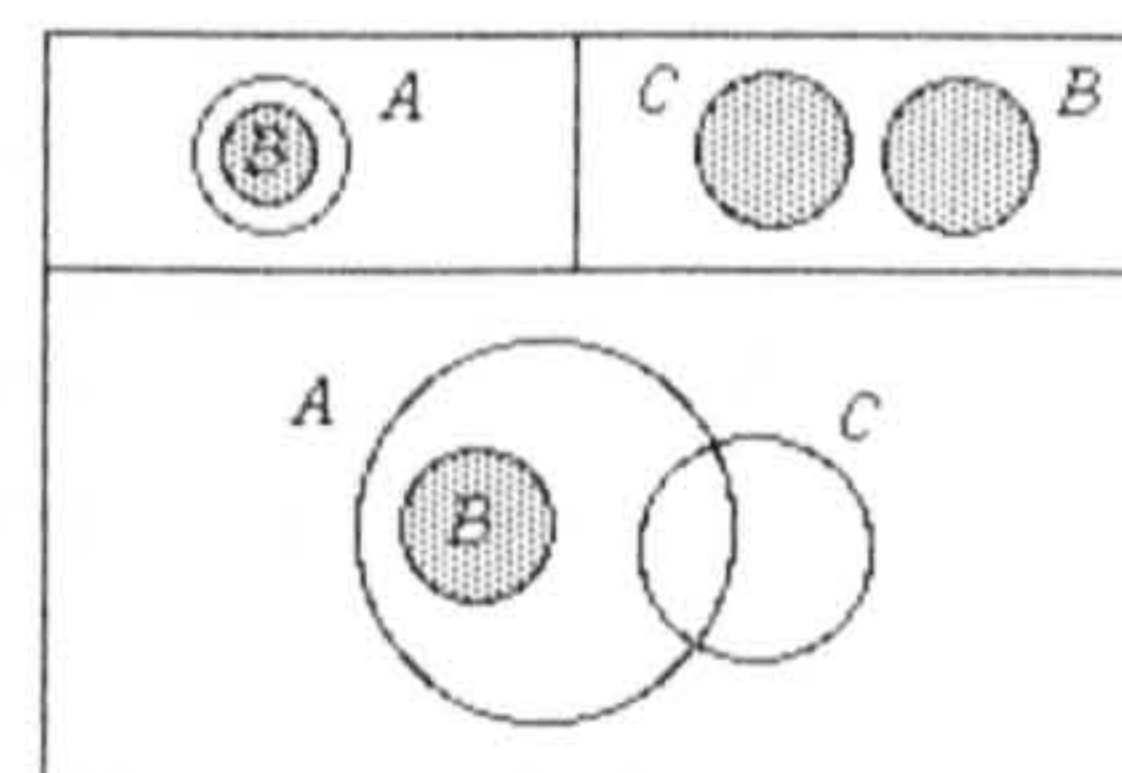


D10

Some subjects seem to miss out stages of the solution process and still get the final solution correct. We think these subjects are taking a short cut to find the critical region and get to the solution.



D11



D12

Both the diagrams D11 and D12 provided by the same subject are correct in the premises. For D11, the problem was, Some Books are fictional stories, No radio play is a book



and the conclusion offered was, Some Fictional Stories are not plays. For D12, the premises were, All Bridges are High, No tunnels are Bridges and Some High things are not Tunnels. The shading is not maintained to the conclusion diagrams indicating a NS type error. However the critical region is shaded in each case showing that the subject has correctly located the critical region but has missed out a number of stages in the process. This behaviour has been listed as NCW, or novel conventions which can work for the subject. Only one of these situations occurred in this study, however it is worth pointing out. Many disciplines use notations to teach abstract skills and concepts. Mathematics, logic, music, linguistics and other subjects all rely on notations. It is likely that especially in the early stages of learning notations, a learner who had grasped a concept but was using a different notation to represent it, would be very discouraged to be corrected without explanation. It is likely that many of the problems learners have with abstract material, are related to the different notations used in different places, and in the failure of the instructor to clearly distinguish between the notation and the abstract principle itself. It would be unfortunate to discourage a learner who had tentatively understood the concept but was, using a different notation to communicate it.

6.3.8 Bug distribution in Euler

There are seven bugs and errors which can be simply determined from the diagrams which learners leave after posttests using the Euler systems. Values represented in Table 24, are raw numbers of bugs from subject data with fractions of bug types as proportions of diagrams generated for particular conditions. In the EULER conditions, 109 diagrams were provided from a potential 210 diagrams.

Table 24: The error spread between learner groups using the EULER system..

	Bug types						
	NS	NPD	IL	TOE	MCI	Path	NCW
M (d = 70)	43 (.61)	38 (.54)	5 (.07)	28 (.40)	7(.10)	3 (.04)	7 (.10)
H (d = 39)	25 (.64)	16 (.41)	0 (.00)	14 (.36)	17 (.43)	0 (.00)	0 (.00)

Note: d = number of diagrams submitted. H = Humanities. M = Mathematics. NS = No Shading, NPD = No Premise Diagrams, IL = Incomplete Labeling, TOE = Premise Term Order Error, MCI = Missing Circle Intersection, Path = Pathological, NSW = Novel Shading Convention which may work.

As for the Venn system results above, there are differences in the kinds of bugs exhibited by the different groups. In the EULER conditions, the difference between the mathematics and humanities conditions is less obvious. High incidence bugs are high for both humanities and mathematics groups in these conditions. There are two interesting results of



this kind. Humanities students had many more problems finding intersections between regions. The MCI bug count for the mathematics group is much lower than it is for humanities students. Overall, mathematics students made significantly fewer mistakes (Mann-Whitney,  $p = .01$ ,  $U = 42.5$ ,  $U' = 6.5$ ). One learner developed an understanding of the system so well that he was able to effectively short cut the learning of the full system and to just shade the critical region. That region was the one which remains nonintersected after registration when there is a valid conclusion. This learner was in the mathematics group.

#### 6.4 Discussion.

The work reported in this chapter initially set out to evaluate the information enforcement principle to see how it might contribute to a modality choice algorithm in a teaching system. The analysis reported in Chapter 4 had provided a method for interpreting this principle numerically. The experimental design was primarily organised to find any correlation between learning outcomes and the results of that analysis.

The first outcome of the study was that this correlation did not occur. The principle did not predict the learnability of the reasoning skills from the systems. The results showed the opposite of what was predicted from the metric, that the least specific system produced the best learning outcomes. The interpretation of the principle is understood to be fair, and was checked several times with its originator (*e.g.*, Stenning, personal communication, August 29, 1996). There are a few minor reservations about the way in which the metric was derived from the principle, however they are not enough to change the rank order of the systems as they are described here. It is more likely that the principle itself, whilst important, does not have such an overriding importance as was thought. Other variables are likely to have an impact on learnability and several of these were not controlled for in these studies.

There are several features of representational systems that may be important beyond the simple specificity ratio. One problem may be in the way in which the metric was calculated. Complexity can be described in different ways. The gross number of states in a system, provides one way to measure that complexity. The gross number of states may have been created from combinations of similar conventions, or the system may have a new convention for each state. In complexity, this difference is described as the *dimensionality* of the system.

Intuitively, it is possible that systems with the same measure of expressivity may have different learnability. If a system is created from a few common and repeated conventions, it may be easier to learn than a system made from several dissimilar conventions. Dimensionality describes the way in which numbers of states are created. A system with ten possible states may have two variables where each variable may be in one of five states. An example system could have two lamps, each with five levels of brightness. Another system

may have ten completely independent states. The difference may be described in terms of dimension, the first system having two dimensions and the second only one.

The background experiences that learners bring to a learning situation will affect several aspects of their learning. Learners always have some prior knowledge and expectations about how things are going to work. The next chapter looks in detail at how the learners expectations affected the way they perceived the meanings of conventions used in the syllogism representations.

The TARSKI'S WORLD system proved to be unhelpful to most of the learners. The information metric has not been applied to the TARSKI'S WORLD system, because it is not a finite state machine. However, this fact alone makes the system much more expressive than any of the other software used in this project. In some ways the fact that TARSKI'S WORLD was so unhelpful, does help to show that the expressivity of a system is related to its benefit to the learner. There are many features which make TARSKI'S WORLD difficult to use. To begin, the opacity of the predicate calculus compared with the relatively simple and familiar language of circles and labels, increases the overhead to the learner. Translation from the natural terms in a problem, like architect, to labels in the syllogism like A, and again to a lower levels of abstraction in the situated blocks world, like Cube, mean three levels of translation, where many errors can be made. The representation in TARSKI'S WORLD is very situated. The Henkin-Hintikka game (see Barwise & Etchemendy, 1991), will only verify truth and not validity of statements. Using TARSKI'S WORLD to prove inferences in the syllogism must involve exhaustive search for situations in which a statement is not true. The leap needed to see why this is necessary is significant. The main problem with TARSKI'S WORLD in supporting syllogistic reasoning is the difference in assumption made in the underlying logic. In TARSKI'S WORLD the Aristotelian no-empty set assumption is dropped, prohibiting the inference from All to Some. Statements of the form All As are Bs for example, All Tetrahedrons are Large, can be true when there are no Tetrahedrons in the world at all. Instructions are included in the training materials for using TARSKI'S WORLD. They are told to always include at least one such individual in the worlds they build that refers to terms in the premises in their created worlds. This extra effort adds significantly to the learners' overhead in using TARSKI'S WORLD this way.

There are many features of TARSKI'S WORLD that make it difficult to use for supporting syllogistic reasoning. The system's high level of expressivity is compounded by the underlying logical assumptions and the situated nature of the microworld. The finding that TARSKI'S WORLD was so difficult to use, is in contrast to findings of Van der Pal (1995), who claims the need for situated representations in tandem with the formal, for learning the simpler logical connective, material implication.

Other work attributes the benefit of graphical systems to their *authenticity* or to the *situatedness* of the representations. It is worth considering if the interpretations of results in



this study would fit those alternative explanations. In his doctoral dissertation, Van der Pal, (1996) used TARSKI'S WORLD to show that formal and situated representations were necessary to learn material implication.

Learners in this study showed problems in understanding the formal languages and translating from the natural problem to a world made in the software. Searching for situations that contradicted a proposed conclusion to a syllogism caused significant difficulty. Finally, learners had problems translating conclusions back into the terms of the original syllogism.

TARSKI'S WORLD is a highly situated environment. A high level of learner activity is required to construct an understanding of the relationships between objects (Jonassen, 1991). A dialogic process exists in the form of the Henkin-Hintikka game (Barwise & Etchemendy, 1991), conforming to the need for negotiation of meaning (Duffy, Lowyck & Jonassen, 1993). The application supports relatively authentic interaction, a small world of objects and relations, and dialogue to support construction of meaning with the learner. However, TARSKI'S WORLD is also a graphical microworld that can be described with the language of graphical representations.

The TARSKI'S WORLD software has been successfully used to learn the more simple principle of material implication. It's more sophisticated sibling, HYPERPROOF, helps certain kinds of students more than traditional syntactic courses (Stenning, Cox & Oberlander, 1995). Resolving these apparently incompatible outcomes requires a certain amount of speculation.

The main difference between the TARSKI'S WORLD software and alternatives lies in the information enforcement. The success of the blocks-world microworlds were greatest for domains with substantially more complex material. The syllogism is a small fragment of logic. Much more can be expressed and learned with the blocks-world programs. While situativity is likely to provide useful guidelines, the results of several studies appear to show that development of an information based approach is still warranted.

The most unexpected result from this study, was the drop in humanities students scores after using the software. Many teaching interactions have no positive effect on outcomes, but few ever manage to decrease learners' performance. It is possible that, although humanities and mathematics learners score similarly well on pretests they may be doing so for different reasons. This difference may be the key to the later changes in performance. Humanities students may be making particular use of belief cues in early stages of problem solving. During exploratory learning with the software and through the instruction, the importance of belief cues may have been undermined. Problems that have believable conclusions turn out to be invalid during this phase.

Believability cues are both useful and counter productive in real life reasoning. Making an unbelievable inference from data taken to be contingently true, will challenge currently held beliefs and may cause an affective reaction much stronger than reaching a conclusion

that does not challenge beliefs (*see for example*, Wasson & Johnson-Laird, 1972; Revlin & Leirer, 1978; Newstead, 1990). The realization that these cues are not necessarily useful in solving syllogisms is a fact demonstrated by both EULER and VENN applications. Learners who find a particular strategy is not generalisable will eventually reject it. However the use of belief cues is often useful in real-world reasoning and even for problems with believable and valid conclusions. Rejecting the use of these cues may leave such learners without a viable strategy. Learners left in this situation would naturally degrade in performance having rejected a partially useful set of rules, without replacing them with better ones. Without more experiment the hypothesis is not supported but this study could have caught these learners without a viable strategy.

The distribution of bugs shown by humanities students appear to support this hypothesis. These bugs occur at very early stages in the problem solving process. This reflects the difference in approach to the systems between the two conditions. Humanities students show an overall lack of understanding about the representational systems. They often fail to realize that the systems are capable of being used as systematic devices that will produce a result after following the dynamic constraints of the system. Because of this they fail to see that systems can demonstrate information which was not obvious from initial statements (input). This is regarded as a misconception rather than a bug or an error. Misconceptions can be thought of as route causes for some buggy-rules. In the EULER conditions, *pathological diagrams* and *externalizations without premises* indicate this misconception, whereas *term order errors* and *diagrams with missing intersections* are more likely to be caused by incomplete models of the process and are therefore buggy-rules. In VENN the same misconception is demonstrated by *representing the conclusion as a premise* and *constructing spurious conventions*, but not particularly by *empty set errors*, *star movement errors* or *third term errors*, which are again considered as buggy-rules.

### 6.5 Design implications

Collating errors in syllogistic reasoning and matching them with recognizable patterns of diagram errors has produced a valuable resource for improving teaching and learning in the field. This kind of enterprise has several benefits for learners and teachers. The BUGGY system (Brown & Burton, 1978) for example, was successful in helping student teachers to understand that learners who produce incorrect solutions are not simply wrong, but operating under their own mistaken or missing rules (Young & O'Shea, 1981). The identification of the errors that learners demonstrate also helps student teachers to better predict the problems they are likely to face in the classroom.

Some research argues there are limits to the value of this kind of approach to learning systems design. Laurillard (1988) argued that identifying error categories should be



abandoned. This position stems from the belief that determining teaching strategies based on error models has two negative effects, (a) a failure to locate the underlying and more far-reaching problems and (b) a perpetuation of the technification of the process (p241). Laurillard claims that misconceptions and mal-rules are best remediated by providing multiple representations that allow translation. The implication is that the rules do not need to be explicitly known by the system or teacher, for this translation to occur successfully.

We know that planning the right conditions for exploratory learning requires substantial guidance (Elsom-Cook, 1990). The following studies show that well designed interactive graphical systems can help learners to confront errors and misconceptions about the domain. The knowledge of those misconceptions and errors is useful to the designer of those interactive graphics. An awareness of the errors can help designers to choose between systems that provide cognitive conflict in the specific areas that learners find most conceptually or procedurally challenging. This may mean that bugs described in this project, are unnecessarily detailed for an exploratory strategy based on translation. However, matching the strengths of representational systems to areas of learner weakness still appears to be very important for design.

The process of finding misconceptions and errors could be improved by using the learner as a rich source of data. Using just the diagrams produced by the learner ignores several layers of depth in those errors that might be better found from working directly with the learner. The study in the next chapter uses a more learner intensive and interactive data collection approach. The method draws from the phenomenographic tradition (Gilbert, Watts & Osborne, 1985) and extends that approach with quantitative measurement.

The studies reported in the next chapter deal with the failure of humanities students to learn syllogistic reasoning from the graphical systems. Differences in the outward behaviour of the two disciplines must correlate with differences in training or styles shown by members of those populations. The study investigates correlations between cognitive styles and discipline preferences.

The limits of the specificity principle are investigated by looking closely at what learners bring to the learning situation. Preconceived understanding of the graphical systems have an impact on the benefit of the graphical systems. The study assesses which of the systems is preconceived more accurately.

## 7 Prior Expectations & Learner Styles

### 7.1 Introduction.

Several interesting results came from the studies reported in Chapter 6. Two of those results led to the work in this chapter. These results represent surprising and unexplained phenomena that called for a more detailed inspection. The information enforcement principle, central to the investigation since the work in Chapter 3, did not predict the learning outcomes in the way expected. The results showed an opposite effect, where systems that were more specific, were significantly worse for students than others. The second outcome this work addresses, is the marked difference in scores achieved by the humanities and mathematics students. The studies described in this chapter address both of these surprising results.

Several factors may have contributed to the failure of the information enforcement principle to predict learning outcomes. Creating the metric from the literature created certain difficulties and errors of interpretation may have been made. The metric was checked with the author of the specificity principle (*e.g.*, Stenning, personal communication, August 29, 1995). A few minor reservations suggested minor improvements, but the metric appeared to embody the principle.

One particular reservation suggested that the denominator in the ratio should be derived from the number of 'models' expressed in the syllogism, rather than the number of possible configurations of the problems. This difference is important, but in the end does not affect the thesis. Models can be the same for different problems, for example, changing the term order in an A-type premise, changes the problem but the model remains the same.

Modifications based on this criticism change the overall ratios by reducing the magnitude of the denominator in each case. Instead of measuring the expression of a system like Venn, against the 64 alternative problem types, each system would be measured against the smaller number of models in the set of syllogisms.

The information enforcement principle describes an aspect of a system's complexity. Complexity theory is a field of mathematics that includes analysis of algorithms (*e.g.*, Bovet & Crescenzi, 1994). Measuring the gross expression of a system could have been done in several other ways. Finding the product of all permutations for all tokens provides an exact measure of the different configurations that may be produced by a system. This measure does not account for differences in the methods used to achieve these numbers of states. Chapters 6 and 8 both discuss ways to improve the method of counting system states.



Although improvements to the metric are possible, factors producing the outcomes in unexpected results of Chapter 6 are not likely to be explained by these improvements.

This chapter reports three somewhat separate studies. The first study addresses the surprising difference between humanities and mathematics students' performance. Chapter 6 describes some possible explanations for these differences. Learners with mathematics backgrounds and those with humanities backgrounds show different bugs and errors (see Tables 22 & 24). This indicates that strategies used by each group may not be the same. Tests of cognitive style are able to distinguish between approaches to learning. The spy-ring test (Pask, 1975) was administered to see if the different strategies conformed to differences in recognised cognitive approaches. This first study is an implementation of a cognitive style test to demonstrate correlation between humanities and mathematics students with the operationalist and comprehension learner styles based on the CASTE and SOLA systems (Pask & Scott, 1975, Arshad & Kelleher, 1993).

From the studies up until this point, it was clear that representations did not simply provide a list of alternatives with calculable information displaying attributes. Elements of each of the systems appeared to cause different reactions from learners. Learners found some conventions predictable and simple to understand. Other conventions were not intuitive. Difficult conventions clearly cause an additional load for the learner. The second study (the card sorting task) assesses learners' difficulty in understanding the conventions used in these systems. An interview with each of the subjects followed the card sorting task. The card-sorting task used representations of syllogism premises in each of the systems. Because the cards do not represent the whole syllogism, the task does not require an understanding of the inference mechanisms of each system. The card sorting task is a test of the static semantics used in the systems. The interview part of this technique stems from the 'interview-about-instances' approach (Gilbert, Watts & Osborne, 1985). The study shows that representational conventions vary in their ease of understanding because of preconceptions learners bring to learning situations.

The third study provides a vehicle for detailed analysis of eight case study learners learning syllogistic reasoning. The instructional dialogue between a teacher and learner, mediated by software and paper based representations, are recorded on video tape. The study used exploratory sequential data analysis (ESDA). The ESDA method is used in human computer interaction research, for example by Harrison (1995). This study adds to the buggy-rule categories developed in Chapter 6, and provides further insight into the scaffolding capabilities of interactive graphics (Brown & Burton, 1978). It also highlights differences between experts and novices in the domain. The study outlines two successful learning activities with time related aspects. These results would have been unnoticed without this form of data collection and analysis.

The order in which the studies were conducted and reported, with the number of subjects in each study, are as follows.

1. A cognitive style test ( $n = 20$ ), looks at differences in learning approaches.
2. A card sorting task & interview ( $n = 20$ ), looks at static semantics, and,
3. Training case studies & exploratory data analysis ( $n = 8$ ), looks at dynamic semantics.

All three studies used a total of twenty subjects in all. All subjects participated in two studies and eight participated in all three studies. These individuals included eight researchers from the Institute of Educational Technology at the Open University and twelve students from a local secondary school. All subjects did both the card sorting task and the cognitive style test. Eight individuals were selected from the card sorting study group who showed very high or very poor correlation with the accepted meaning of the diagrams used in the cards. The four highest and four lowest scoring subjects from the card sorting task were chosen for the training case studies. The subjects are represented in Table 29. These subjects were then taught the syllogistic reasoning, half with software and half with pencil and paper versions of the systems. The teaching and learning interaction was videotaped and analysed using qualitative analysis approaches, including the Timelines software (Hartson & Grey, 1992; Harrison, 1995; Mackay, 1989). The teaching interaction was supported by an expert tutor. Teaching materials used, were based on those from the study reported in Chapter 6. The VENN and Edinburgh-EULER software applications were used in software supported conditions.

## *7.2 Cognitive Styles.*

The decline in the scores of humanities students after using the software reported in Chapter 6 was a strange outcome. Experimental teaching interventions often yield no significant improvement. It is much more unusual to see significant decline in performance as a result of a teaching strategy. The circumstances for a decline in performance like this are likely to involve a conspicuous change in approach during the learning period.

The distribution of errors for both VENN and EULER circle systems in the specificity evaluation study (Chapter 6), were quite different for each group. The first group showed more elementary errors than mathematics students.

The humanities condition bugs include an aversion to the systematic nature of the representations. This manifested in the idea that new conventions could be arbitrarily invented and that the representations were for annotation and were not generative in any way. Instead of representing the premises and reading the conclusion from the diagram, these students took an offered conclusion and entered it in the diagram as if it were another premise.



By contrast, mathematics students would fail to see that, once two premises had been annotated, if a single region was left in a Venn circle, then a star could be added and an existential conclusion could be made. Other errors like these are more advanced and demonstrate that learners understand the systematic nature of the representations, even if some of the details are not properly practiced.

The pilot study reported in Chapter 5 and the evaluation of the specificity principle in Chapter 6, collected little data relating to the effects of beliefs on learners' performance. The pilot study used only three students. Each of these students were actively involved in advanced level mathematics. The study reported in Chapter 6 carefully controlled for the effects of belief on reasoning performance. Even the card sorting task did not contribute to understanding about how beliefs effect performance. However, both the pilot study and the literature review in Chapter 3, clearly show that beliefs are very important in interpreting premises, processing problems and in evaluating the validity of conclusions.

This study draws on an hypothesis that helps explain the failure of non-mathematics trained learners at learning syllogistic reasoning from the circle systems. The hypothesis claims that learners in each group treat the content and the form of the problems quite differently. Learners with mathematical training are more indifferent to the effects of empirically untrue premises. Whereas learners with different training tend to make use of belief related cues to help evaluate their reasoning process.

Using the yardstick of believability to assess results of every day reasoning problems is an effective strategy. However, the circle systems are able to undermine confidence in this benchmark by showing that empirically untrue statements can result from both untrue premises as well as true premises. The hypothesis is that the non-mathematically trained learners drop the unhelpful use of semantic cues and, for a time at least, have nothing to replace it. The learners founder with no strategy and score no more than chance success on tests of the reasoning skill.

The instrument used in the study is a cognitive style test created by Gordon Pask in the early 1970's (Pask & Scott, 1975). The test investigates differences between learners from two different academic streams. The style test directly assesses learners use of semantic information in problem solving. The test effectively establishes which learners make use of the belief related cues around them, and which are less affected by them.

Cognitive style discrimination can contribute to improved teaching (Kolb, 1984; Schmeck, 1988; Pask & Scott, 1975) and the adaptivity of computer assisted tutoring (Arcand, 1994; Dobson, 1996; Arshad & Kelleher, 1993). Learning styles are tracked and assessed from patterns of behaviour. Some learners naturally begin with an overall picture of the domain and then study the details. Others begin with simple steps and accumulate competence towards a full model. Pask's discrimination between serialist and holist learners and between comprehension and operational learners is tested with the 'spy ring test' or by

its virtual isomorph, 'the smugglers test'. Both tests rely on a central idea about the representation of the domain, the entailment mesh. The spy ring test (Pask & Scott, 1975) was selected to deepen understanding of the surprising failure of humanities learners in the previous study.

According to Pask and Scott (1975) the test is designed to capture the differences in learner styles used by individual learners. Learners operating with a 'serialist' style, tend to work analytically by mechanistic memorisation techniques and by encoding information without reference to the context and embeddedness of the domain. Learners with a 'comprehension' style tend to incorporate new materials while making semantic links with the meaning inherent in the domain. They pick up relevant cues from all available information. A third style represents the synthesis of these two alternatives. Learners with this style are known as 'versatile'. These learners are able to apply the appropriate style when the learning problem demands it.

Subjects taking the spy ring test must understand and remember the activities a network of communicating agents during a period of three years. Certain rules limit the kinds of communication possible between each agent. Agents' safety, position and power in an hierarchy, as well as their geographical locations all determine what messages are allowable. The subject taking this test is first shown how to derive directed graph diagrams from lists of predicated relations referring to the network. The subject then has to learn the activities of each of the spies over the three year time period. Along with these formal descriptions of communications between agents, the learners also have material about each of the countries including the terrain, style of government, and national characteristics.

Certain information is implicit in the materials, for example agents may either pass messages but not originate them, or may have to wait for two or more messages before achieving the right to transmit. The test subject must infer the status of the agent from their activities together with the national boundaries that separate the three countries. Figure 19 illustrates a complete history of the communication between agents over a three-year period.

There are no messages passed between Byron and Caesar during the three year period and so the subject should infer from this that Byron and Caesar work in countries without a common boundary. The diagram in Figure 20 shows a possible geographical solution to the spy-ring test. The agent named Ajax is an 'accumulator' type agent. He must receive at least two messages from other agents before he is permitted to transmit. In years 1985, 1986 and 1987 Ajax fits this pattern: in 1985 he receives two messages from Byron before transmitting to Euclid. In 1986 he receives two messages from Byron before replying back to Byron. In 1987 he receives from both Byron and Euclid before transmitting to Byron. This kind of inference is inductive and always open to review. A transmitter-type agent does not have to send messages as soon as they are received. An agent behaving like an accumulator may



eventually turn out to be a transmitter. Once an agent has been identified as a transmitter he cannot become and accumulator.

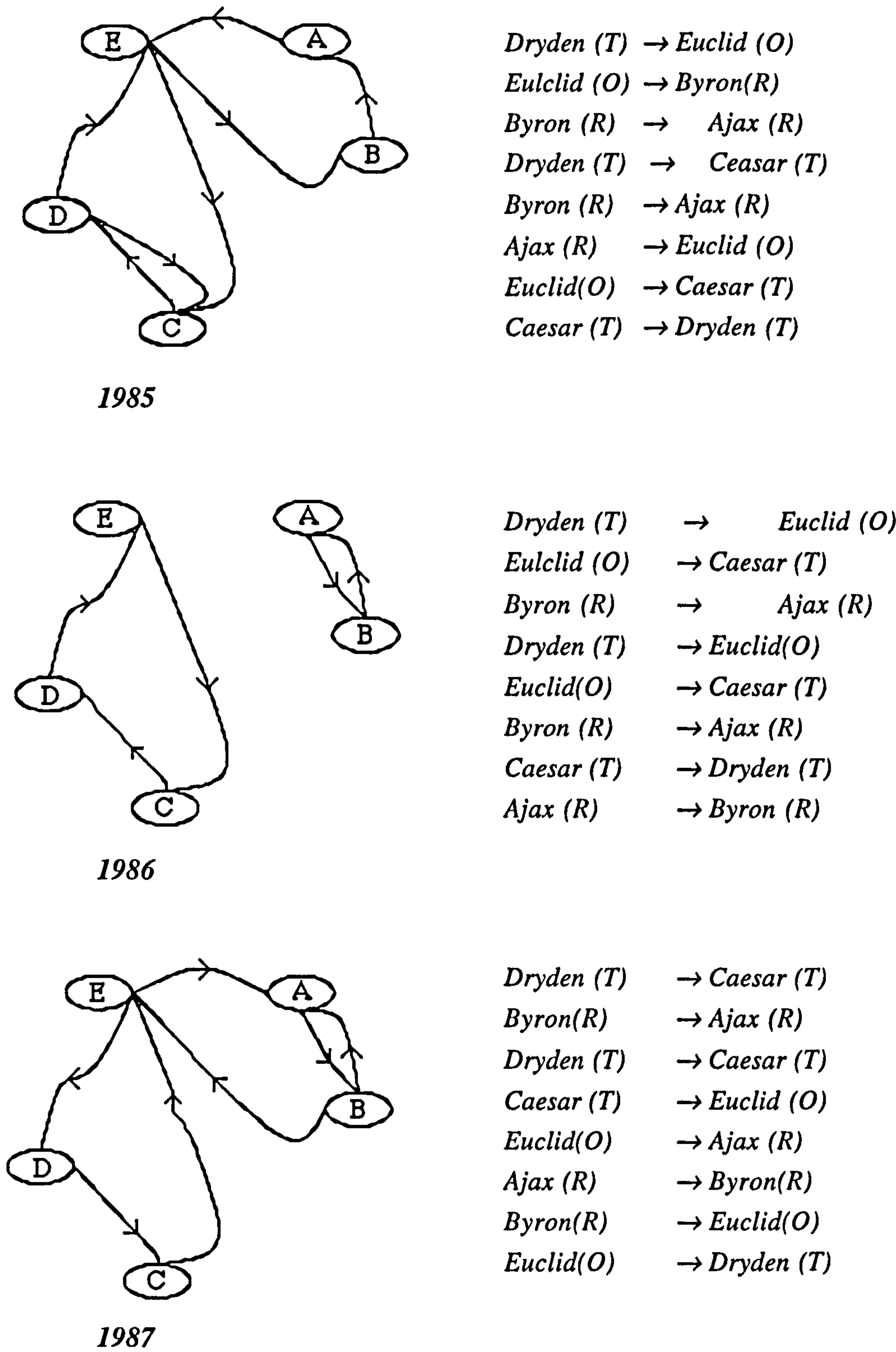


Figure 19: Messages between agents as graphs over 3 years.

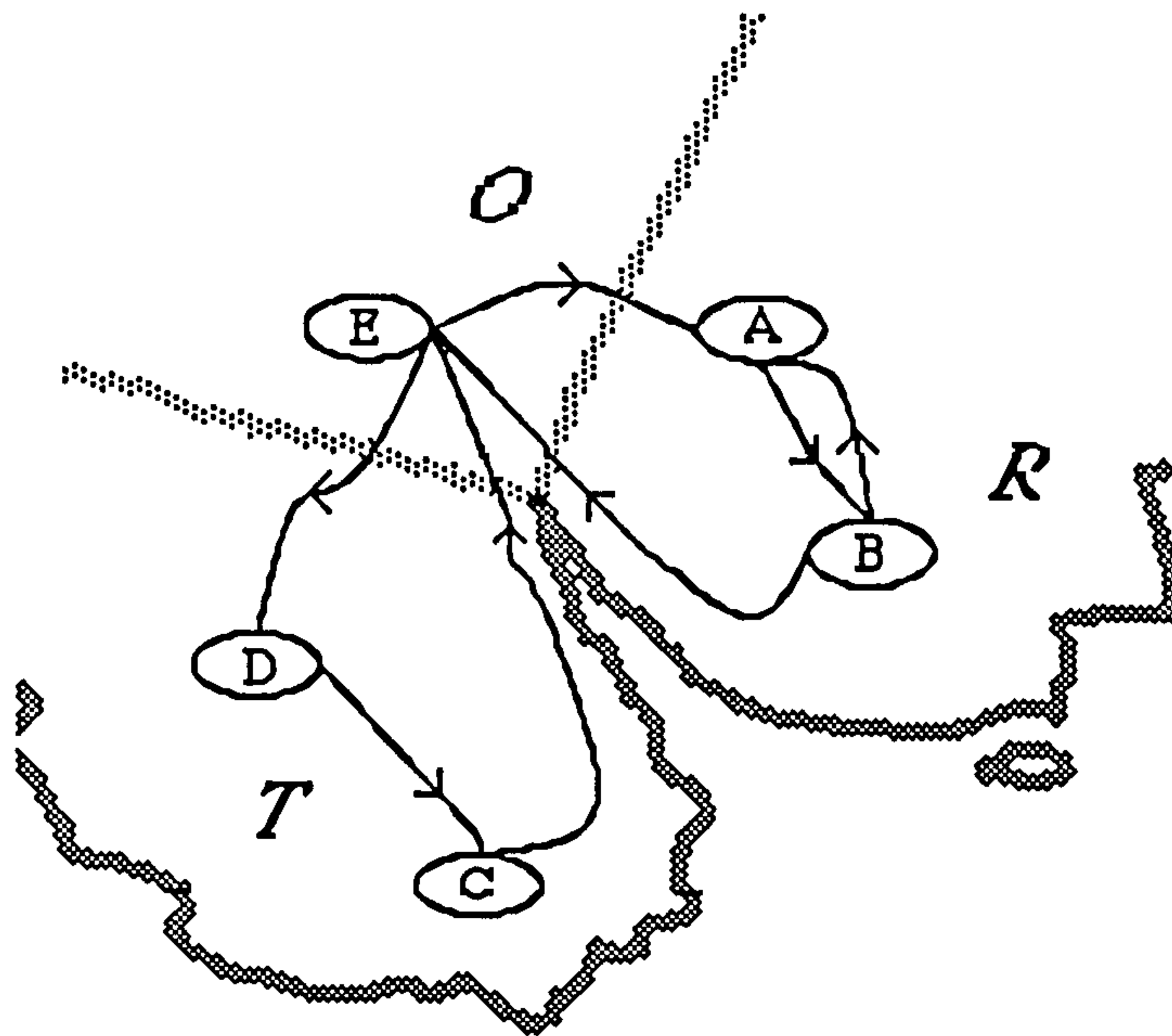


Figure 20: A geographical solution for the spy ring.

The most challenging task for the subject is to predict agent's communication patterns in a year immediately following those described in the materials. This prediction is possible when the learner understands the rules behind the patterns of communication leading up to the last year of messages.

The test also involves many self reported strategy assessment items. These add evidence to the categorization of each learner style. The results are presented as a rating for each subject in four categories of style, (a) operationalist, (b) comprehension, (c) neutral and (d) versatile styles of learning.

The spy ring test was administered individually to the 20 subjects used in this study. The test is a labour intensive instrument and varies in the time required to complete depending on the time taken by learners to remember the material. In the group process this can take a number of hours since each subject must reach the same stage before the group can continue.

### 7.2.1 Subject selection

Twenty subjects completed the spy-ring test. This group consisted of twelve students from a local secondary school and eight researchers from the Institute of Educational Technology. Subjects were assessed for their mathematical experience and chosen field of



study. Current enrollment in, or previous success in advanced level mathematics classes in secondary school, provided the primary criterion for assignment to the mathematics condition. Subjects from the secondary school group were matched for previous scores in practice advanced level exams. Six subjects were not enrolled in mathematics classes but had completed general certificate courses in mathematics. The remaining six secondary school learners were active participants in the advanced level mathematics course at the school. Of the eight researchers four had completed advanced mathematics training. These researchers came from three different native countries with different languages and school systems. Four of the eight explained high levels of mathematics training in their secondary systems. The remaining four researchers had not taken advanced mathematics courses while in the secondary system.

The mixture of researchers and school learners in the design enabled a comparison of beginning learners and competent practitioners with similar mathematical training. The spy ring test did not show any differences in style, attributable to differences in the maturity within each group. Subjects within the groups are homogeneous in terms of mathematical training.

7.2.2 Results

The results in Table 25 show the mean score values for subjects in each of the operationalist, comprehension, neutral and versatile indices. Each value is derived from the ten subjects in each category. Since the total possible scores for each category are not equal, the percentage means are also shown. The full data appears in the appendix at the end of the report. There is a clear difference between the two sample populations. The mathematics trained learners score more highly in operationalist and versatile styles. Humanities learners scores more highly in comprehension style.

Table 25: Results of spy-ring test.

Subject	Learner Style Index							
	Neutral (2)		Operational (7)		Comp (14)		Versatile (5)	
	MS	%	MS	%	MS	%	MS	%
Mathematics	1.20	0.60	4.96	0.71	3.74	0.27	1.62	0.32
Humanities	0.32	0.16	2.77	0.28	6.11	0.44	0.59	0.12

Note : Numbers in brackets indicate maximum score for each index. MS = Mean score over all 10 subjects in each condition.

There is a significant difference between the mathematics and humanities groups on the operationalist scale (Mann-Whitney *U* test,  $U = 1.50$ ,  $p < .001$ ). There is also a correlation between serialist learning styles and mathematics students reinforcing the hypothesis that

mathematics students are less affected by semantic content in problems. Difference of comprehension type learners is only slightly less significant ( $p < .005$ , Mann-Whitney  $U$  test,  $U = 83$ ), and is likely only less significant because mathematics learners also score highly in the versatility category. The difference between versatility scores between learners with mathematics and humanities backgrounds, is also significant ( $p < .001$ , Mann-Whitney  $U$  test). Mathematics students score highly in both versatility and operationalist approaches.

This study administered the spy ring test for learner styles with all twenty of the participants used in the combined studies. The study was designed to explain differences between humanities and mathematics learners. These findings show a very strong correlation between serialist learning as identified by the test and learners with strong mathematics backgrounds as identified by the criteria for group allocation. There is also a weaker correlation between learners identified by the test as comprehension learners and those fitting the 'humanities' criteria for subject group allocation in the design.

We still do not know why comprehension or humanities students' scores drop after intervention with the teaching method, but the study indicates that there is strong reason to believe their decline in performance is at least paralleled by their style of learning. This may indicate that there are appropriate styles for different subject areas, as Pask and Scott (1975) claim.

The mathematics students showed evidence of a versatile approach, as well as highly serialist style. The inability of humanities students to change to this style may well have caused their difficulties with the material. The drop in scores may be attributed to the rejection of strategies that were originally partly successful. Rejecting the partly useful strategy may leave the learner with no strategy to replace it. This hypothesis and the findings of the previous studies in this chapter are treated in the final discussion.

### *7.3 The card sorting task.*

The card sorting task was designed to determine variations in perception of conventions used in the graphical systems to communicate the syllogistic reasoning domain. Learners clearly bring experiences and expectations to the learning environment. The study in Chapter 6 showed that some conventions in the graphical systems provided challenges more than others. This card sorting study assesses which of the conventions are most like learners' expectations and which are counter-intuitive.

This study used the same twenty subjects as in the test of cognitive style. In each case the card sorting task was administered individually and some time after the test of cognitive style. These studies are not strictly independent, since the same subjects were used in each case. However, the nature of the task in each treatment is quite different and there is little reason to suspect any interaction. Unlike the previous study, this one is not intended to find



differences between learners with different backgrounds or learning styles. In this case a more comprehensive sample of the population is more useful. The subjects represent a cross section of the more general population without splitting them based on their background or study interests. Two of the researcher group had been trained in introductory logic and one of these had used circle based diagram systems in their training. This individual was tracked separately through the study.

The study uncovers the meanings of graphical representations that learners assign to graphical representations before they receive any training. The hypothesis is that the differences between the expected meaning (those assigned by the learner before training), and the accepted meaning (what is eventually taught) will determine the amount of work needed to overcome mis-preconceptions. The study methodology is able to determine the rank order of various graphical conventions by using a new addition to card-sorting procedures. This is described in detail below.

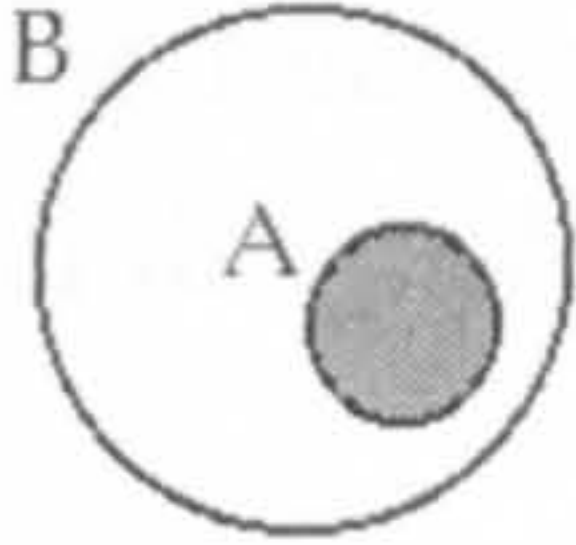
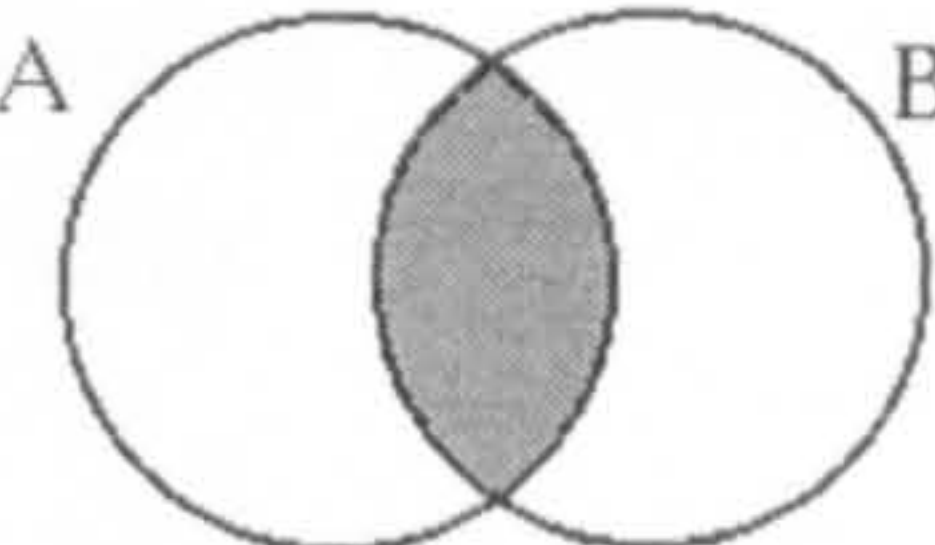
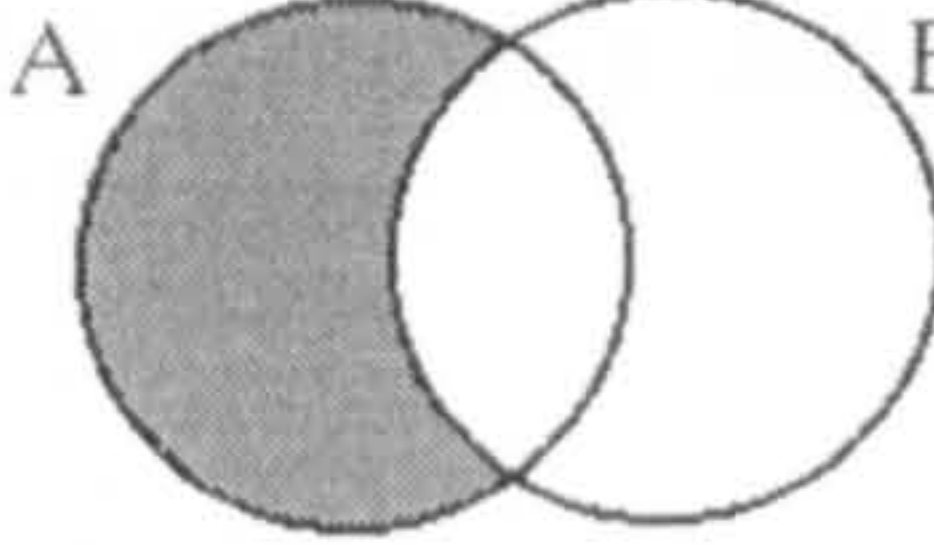
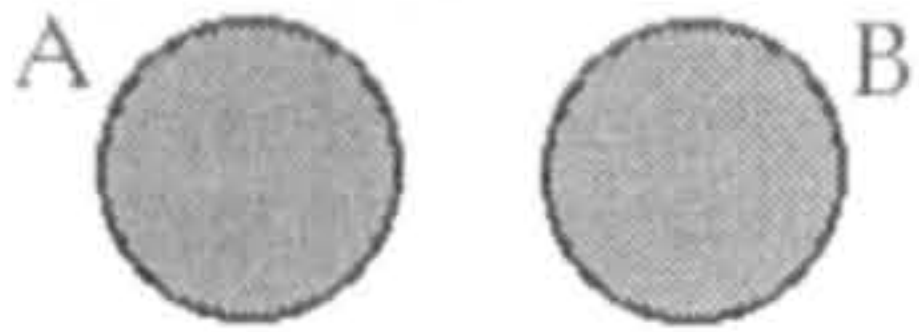
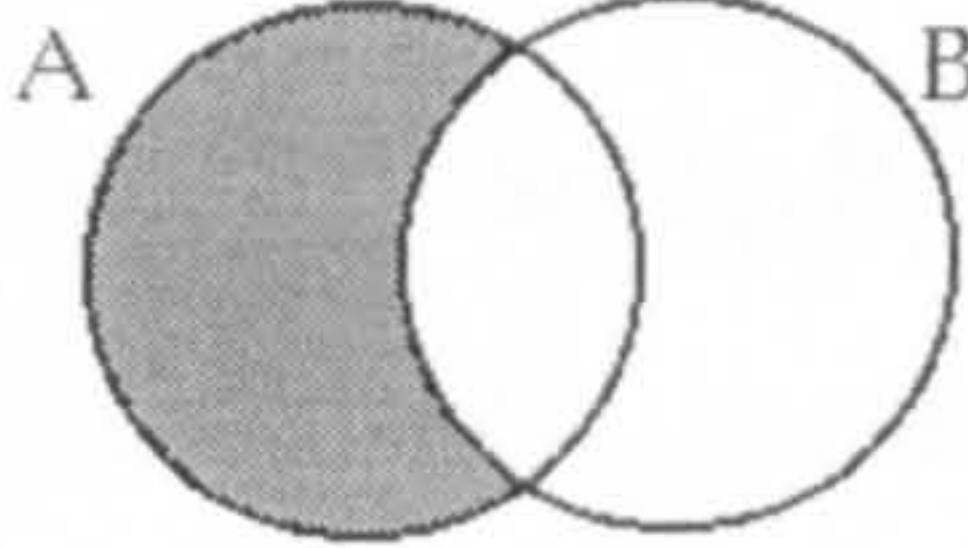
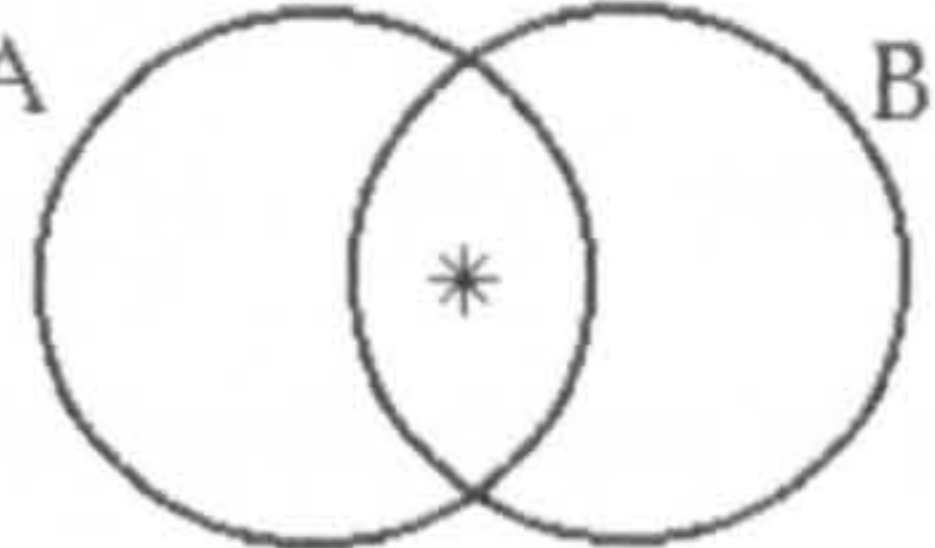
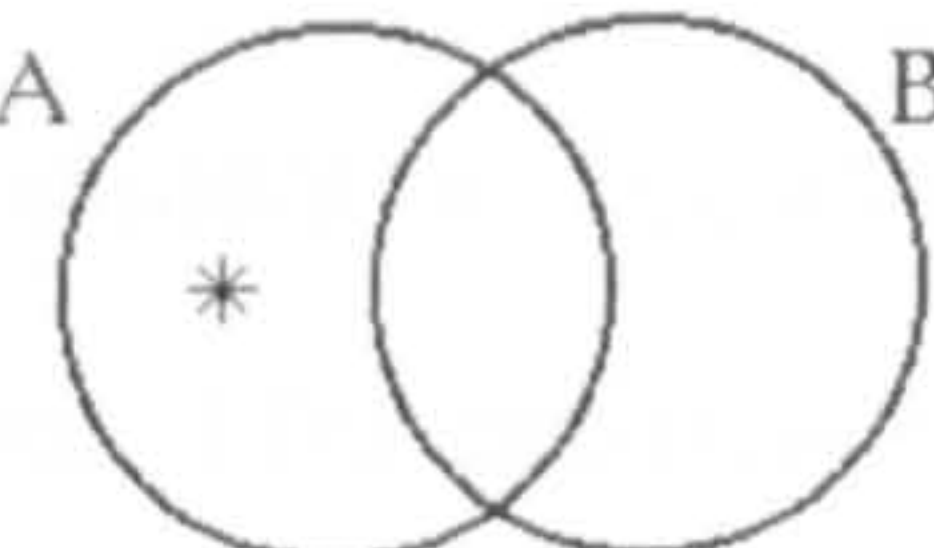
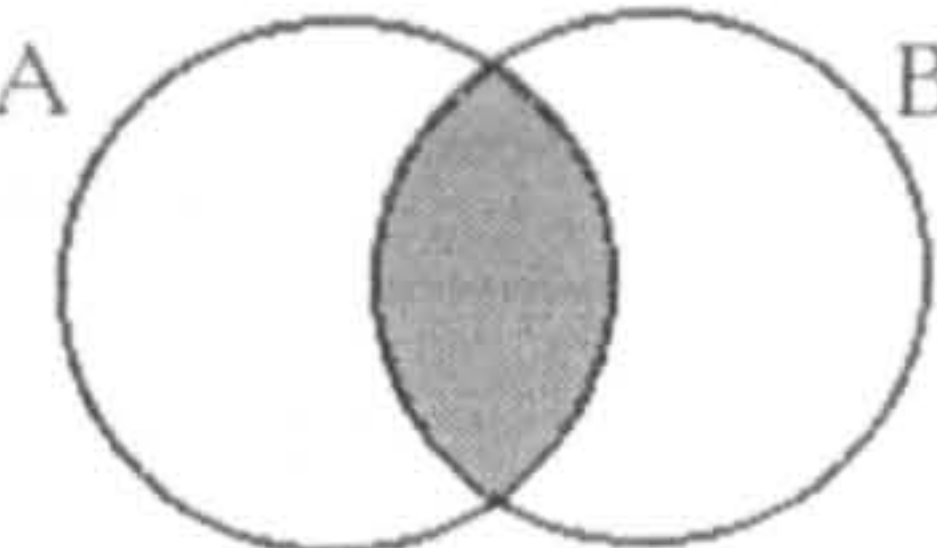
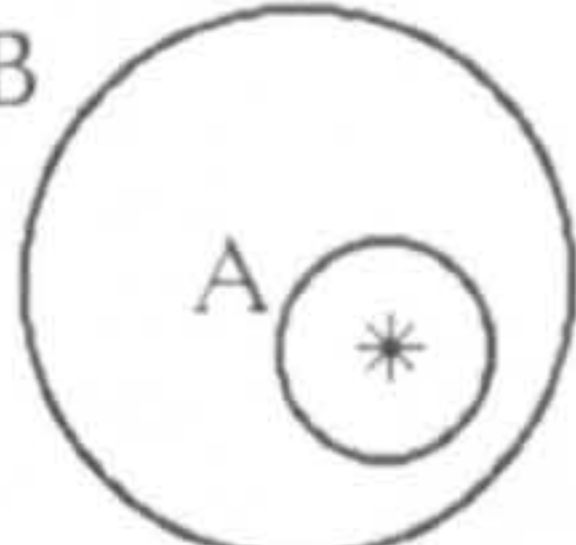
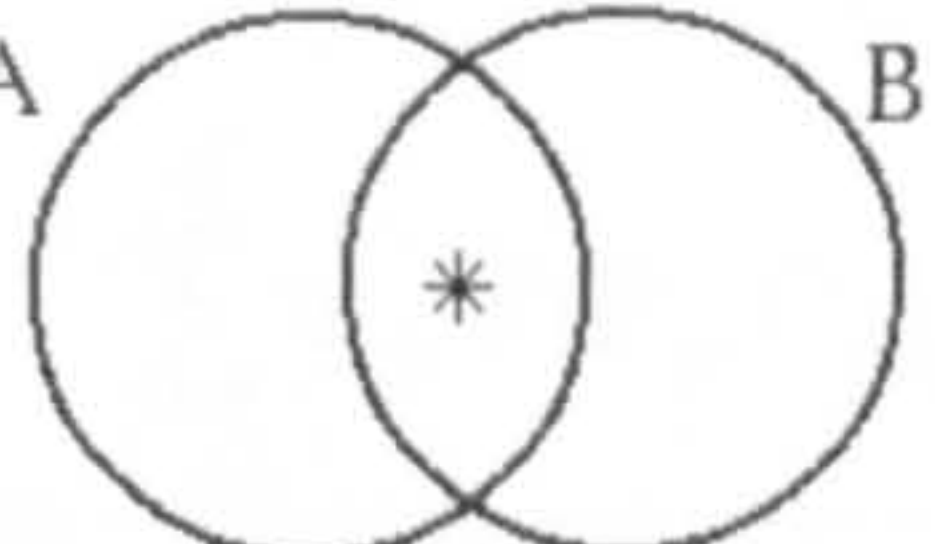
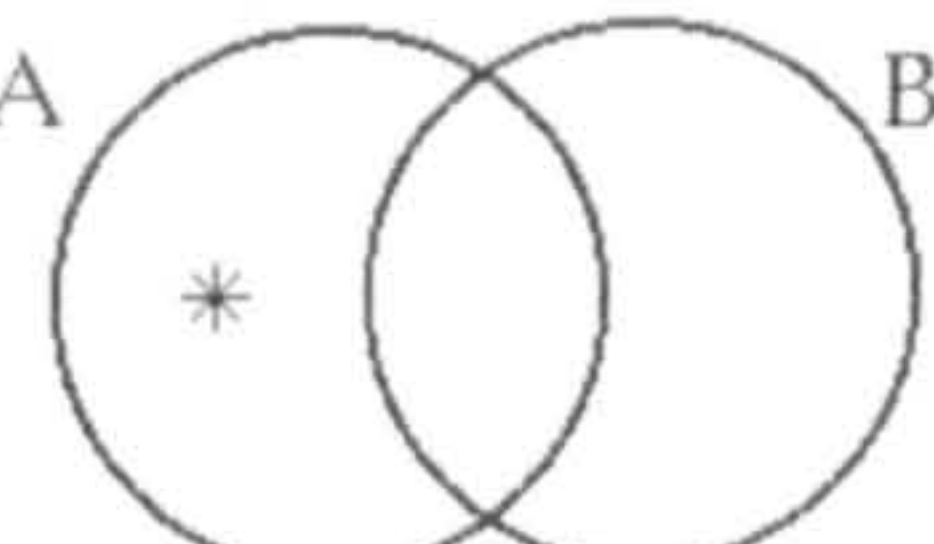

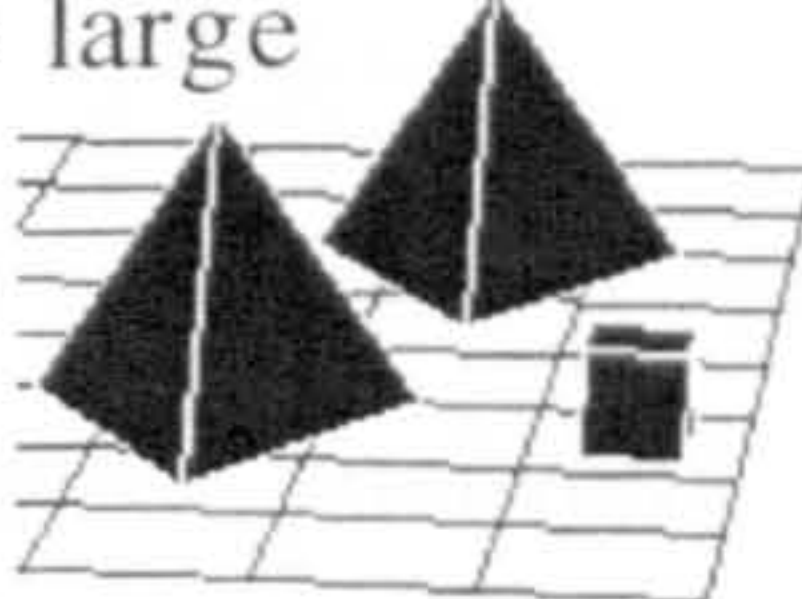
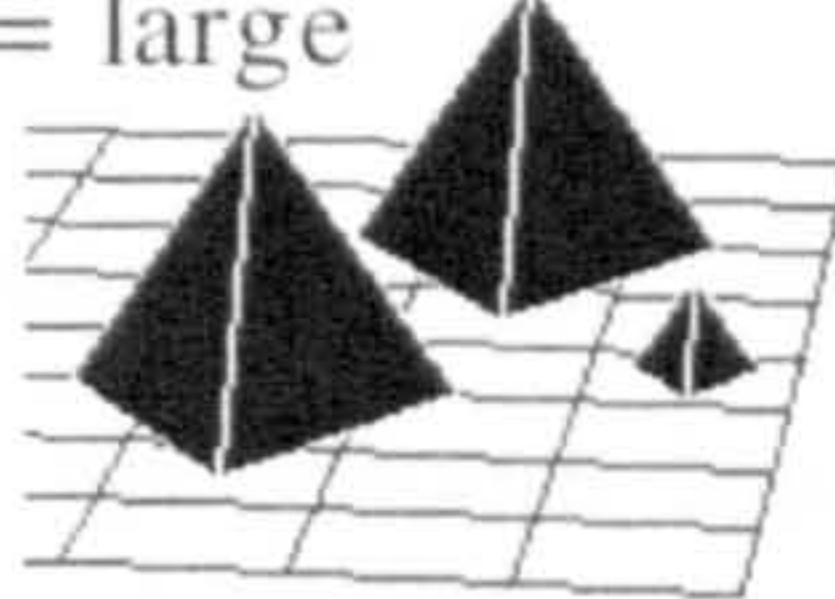
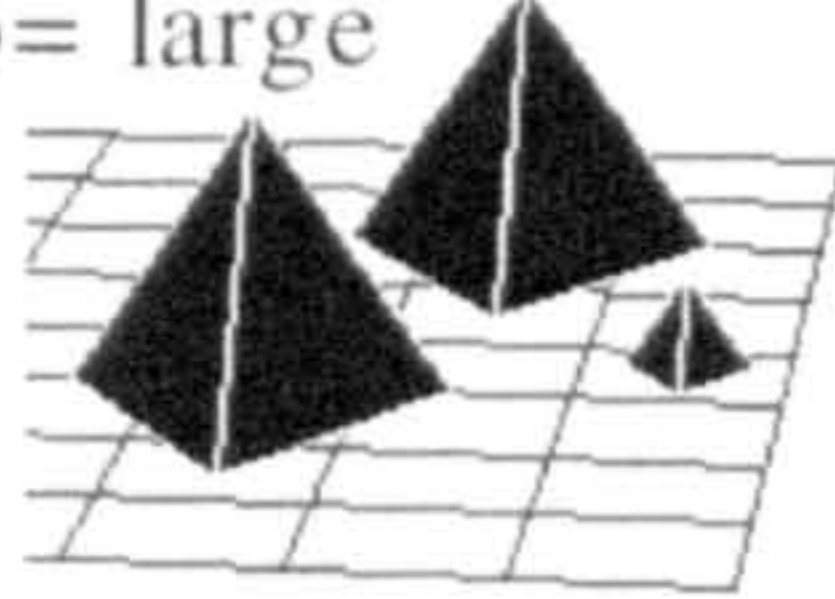

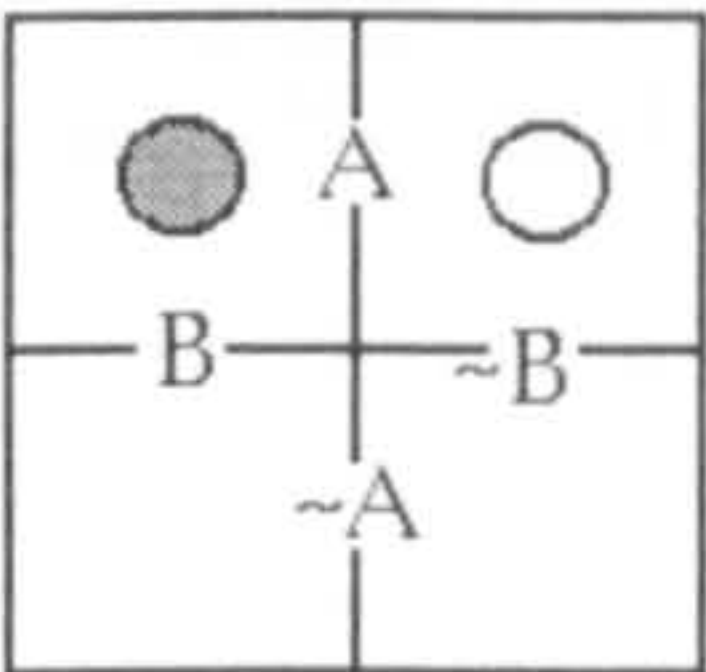
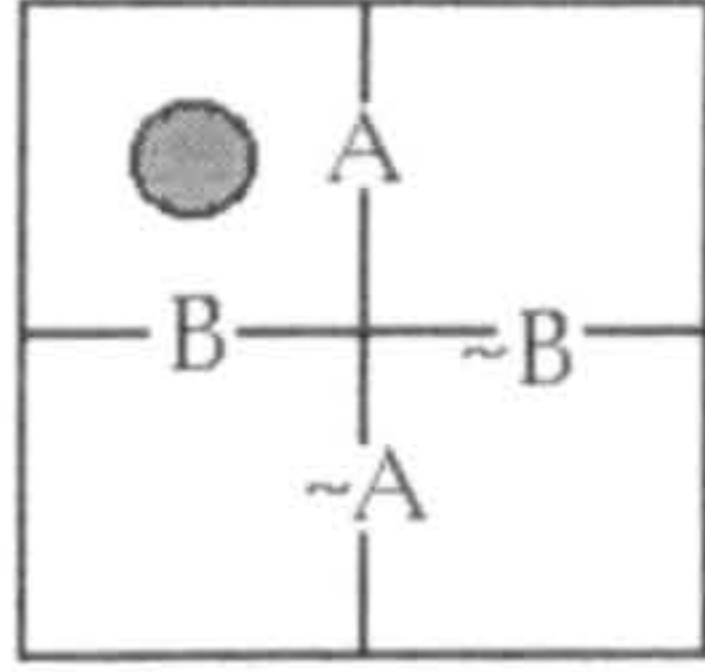
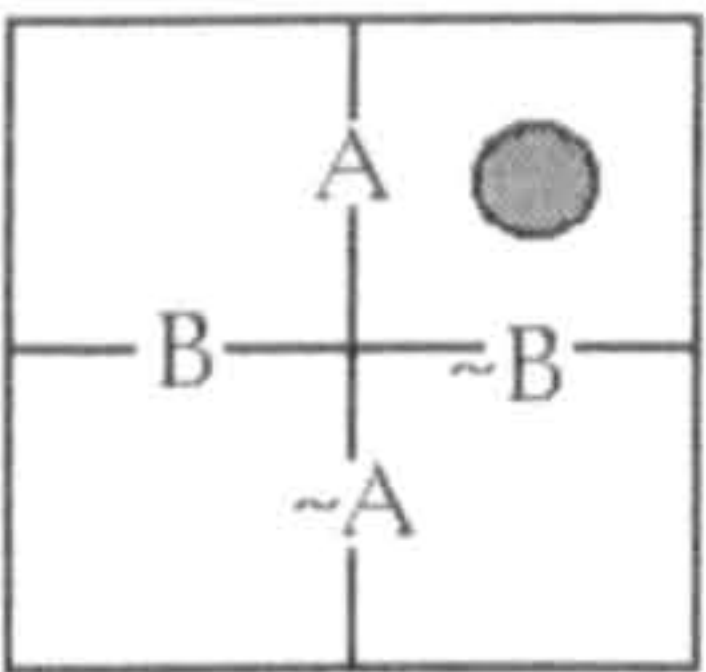
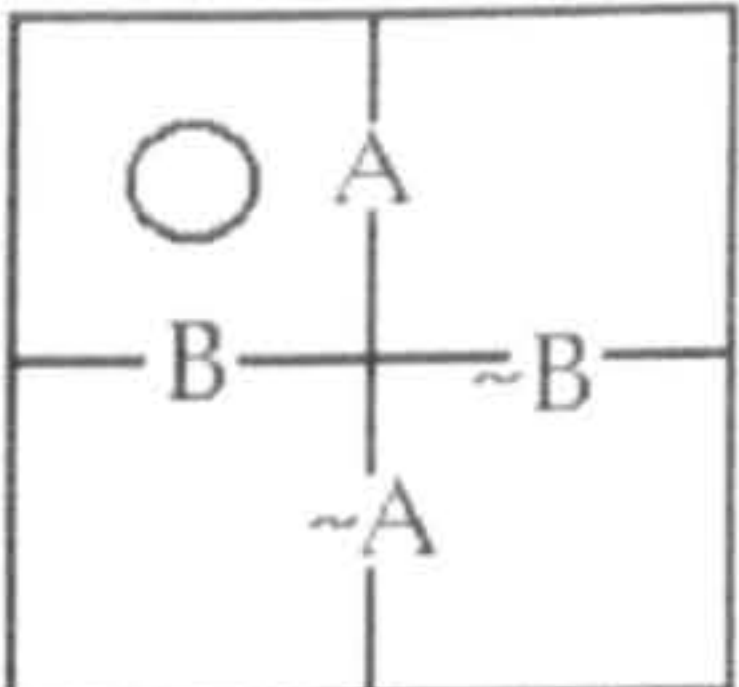
The cards used in the study include one for each of the premises from each of the systems. Table 26 shows these cards. Each of the twenty cards has a representation of a single premise from each of the systems: Euler, Venn, Carroll, TARSKI’S WORLD and Euler\*. The Euler\* system is similar to Edinburgh-Euler but uses a \* to indicate existence instead of shading. The TARSKI’S WORLD cards require labeling as well as the diagrams. The objects in the TARSKI’S WORLD illustrations do not represent the terms used in a premise. The objects are inhabitants of a world described by the terms included in the labels.

The cards represent only the premises of the syllogism. Much of learners’ difficulty, as demonstrated in the previous studies, lies in combining (registering) the premise diagrams. The process for combining these diagrams is often complex. The card sorting study does not address difficulties associated with the registration process. The card sorting test is a method for assessing difficulty in perceiving meanings for the premises and not for difficulties in combining premise diagrams. The following training case studies illustrate some of the problems learners have with the more dynamic aspects of the diagram system designs.

The subjects’ task in this study was to identify the cards from the pack illustrated in Table 26, which they considered best illustrated each of the four premises. Subjects were presented with an empty table with five columns and four rows. A premise type was written to the left of each row in the, “Some As are Bs” format. No semantic information was provided as contextual clues in the premises. Subjects were asked to assemble the twenty cards in order of preference from right to left for each of the premise types. The results of each subject’s choices are managed with the formulae described below.

A second phase of the card sorting study provided a better chance for explanation of the choices made by subjects. Subjects were interviewed with their choices in front of them to enable a post-hoc explanation for their choices.

Table 26: Cards used in the card-sort.

	Premise Mood			
	A	I	O	E
<i>E</i>				
<i>V</i>				
<i>E*</i>				
<i>T</i>	a= tetrahedron b= large 	a= tetrahedron b= large 	a= tetrahedron b= large 	a= tetrahedron b= large 
<i>C</i>				

Note: T= TARSKI'S WORLD, V= Venn, E\* = Euler with shading replaced by the star, C = Carroll. In the TARSKI'S WORLD representations, a = tetrahedra and b = large.

### 7.3.1 Data in the card-sorting study

Scoring the card sorting data required addition to techniques previously described for card sorting studies (*e.g.*, Gilbert, Watts & Osborne, 1985). The subject can place a correct



diagram in one of four positions of preference. To capture the preferences for the diagrams, the positions are given a weighting, from four to one in order of decreasing preference. First choice diagrams receive four points, second place three points, third two points and least preferred, one point. The maximum score that a subject could achieve for a single premise would occur if a correct diagram was placed in each position. This would sum to ten points. The following functions for subject's scores, premise intuitivity and system intuitivity are all created from the card sorting data.

Let,  $w_i = 5 - i \Rightarrow w_1 = 4, w_2 = 3, w_3 = 2, w_4 = 1,$

$$\sum_{i=1}^4 w_i = 10$$

Let,  $Sc(S_i, P_j) =$  score of student  $i$ ,  $S_i$ , on predicate  $j$ ,  $P_j$ .

Let,  $C_{ijk}$  be the *chosen diagram* of student  $i$  for predicate  $j$  for position  $k$

Then,  $V(P_j, C_{ijk}) = 0$ , if  $C_{ijk}$  is wrong for  $P_j$   
 $= 1$ , if  $C_{ijk}$  is right for  $P_j$

So, 
$$Sc(S_i, P_j) = \frac{\sum_{k=1}^4 w_k V(P_j, C_{ijk})}{\sum_{k=1}^4 w_k}$$

For example, a subject correctly matching two diagrams for a mood, in their first and second choice positions, would get a score based on the points for each position, divided by the total possible points.

$$Sc = \frac{4+3}{10} = 0.7$$

This measures how close the subjects' interpreted meanings for diagrams are to the accepted meanings in each of the five diagram systems. It is a measure of how intuitive the accepted diagrams are for a single premise type.

A more general measure of "mood intuitivity," is the mean score for all subjects, for a single premise type. The single subject data is aggregated to show the mood intuitivity for all subjects. This is a measure of difficulty in interpreting each premise mood for the whole population, with all systems. These scores, one for each premise, are shown in the right most column of Table 26. The measure is calculated by the following definitions and equations.

There are four premise moods. The  $\delta$  function picks out the learner's answers that are in the right mood.

$$\begin{aligned} \text{Let, } M(P_k) &= \text{mood of premise } k, P_k = \text{one of } m_i, i=1...4 \\ \delta(m_i, m_j) &= 1 \text{ if } m_i = m_j \\ &= 0 \text{ if } m_i \neq m_j \end{aligned}$$

The premise intuitivity score is then calculated by adding all the weighted scores for correct diagrams, for a single mood and dividing by the number of possible diagrams in that category.

$$PI_m = \frac{\sum_i \sum_j Sc(S_i, P_j), \delta(m_i, M(P_j))}{\sum_i \sum_j \delta(m_i, M(P_j))}$$

To establish the extent that prior expectations affect learnability of different systems the comparison must be extended to contrast different systems rather than different premise types.

The score measured by  $SI_{ms}$  below, measures how close the intended meaning of the system is to the representations subjects prefer for the systems. System intuitivity is an aggregated score for each system, consisting of numbers of correct diagrams in preference positions for all subjects. A single subject, putting the correct Euler\* premise diagram against an A type premise in the card sorting, would score four out of four possible points for the Euler\* system. The same subject, placing the correct Venn diagram in second position, would score three from four possible points for Venn. The maximum score for any single subject and for any system is four points. When these scores are aggregated over all 20 subjects, we get the scores in the Table 27 marked as SMI (for system mood intuitivity).

These scores are the mean scores for all subjects for a system representation for a single premise. When the scores for each of the premises are averaged over the number of systems, the result indicates the overall intuitivity of the particular system.

This measure is important in establishing the hypothesis for the card sorting study. The study investigates how learnability of a system is affected by learners' preconceptions of diagram meaning, independent of the system's computational (information enforcement) characteristics. The values for each system are represented in the lowest row of Table 27. These scores are calculated by the following expression.



Let        Sys(C<sub>ijk</sub>)        =        system of choice  
             δ(s<sub>i</sub>,s<sub>j</sub>)        =        1 if s<sub>i</sub>= s<sub>j</sub>  
                                 =        0 if s<sub>i</sub>≠ s<sub>j</sub>

$$SI_{ms} = \frac{\sum_i \sum_j \sum_k w_k V(P_j, C_{ijk}) \delta(m_i, M(P_j)) \delta(s_i, Sys(C_{ijk}))}{\sum_i \sum_j \sum_k \delta(m_i, M(P_j)) \delta(s_i, Sys(C_{ijk}))}$$

The next section presents results of the card sorting task.

*Results of card sorting task*

Results show a substantial preference for the Euler\* system particularly over TARSKI'S WORLD and Carroll. This may indicate the general preference for representations not using any shading. Subjects in the previous studies thought of shading as an unnecessary highlighting annotation. This phenomenon was indicated in Chapter 6 when learners using Euler omitted shading in over 60% of diagrams. This preference is upheld for A, I and E-type premises over each other system. However the Euler system is preferred for the O-type premise. The Venn representation is the same as Euler\* on E and I-type premises respectively and so values for preferences are the same for each.

As indicated earlier, TARSKI'S WORLD was useful for very few subjects. There seem to be extreme difficulties for subjects using this system, however for those for whom it was useful it proved very unambiguous. Unfortunately, although there were signs of this kind of experience during observation with the learners, these subjects do not show up in the quantitative data.

Table 27: Results from Card-sorting task.

	System					
	Euler*	Euler	Carroll	Venn	TW	PI
(A) All As are Bs	3.00	2.41	0.32	1.50	0.30	1.51
(I) Some As are Bs	2.90	2.10	0.70	2.90	0.20	1.76
(O) Some As are not Bs	2.50	2.60	0.60	2.50	0.10	1.66
(E) No As are Bs	2.20	1.30	0.10	0.10	0.00	0.74
SI (mean scores)	2.65	2.10	0.43	1.75	0.15	

*Note:* Euler \* = Euler system with \* used as replacement for shading. SI = System Intuitivity. PI = Population Mood Intuitivity. Cell contents are SMI = System Mood Intuitivity.

Lower scores on Euler for E and I-type premises, than for A and O-types may be due to competition from Carroll system, for those two premises only. For those two premises the Carroll system is relatively simple for subjects who treat shaded regions as indicating existence of individuals. The A and O-type premise representations in Carroll are not easily understood because of confusion over shaded and unshaded counters. Both these seem to be treated as all indicating existence (see interview below). The shading is therefore treated as irrelevant by these subjects.

These results show that the meanings which learners (before training) attribute to diagrams, are nearer to the accepted meanings of the Euler system than any of the other systems. Learners' preconceptions of the Euler system are more accurate than any of the other systems. In this respect the Euler system should present the fewest problems for learners since they have less work to do than for any other system.

Learners prefer the \* convention to the standard Edinburgh-Euler system, but they prefer both Euler systems more than Venn, Carroll or TARSKI'S WORLD. The star convention is a small modification to the Euler system that improves its acceptance to the user group.

The study investigated a variable that appeared to confound the predictive powers of the specificity principle. The experimental hypothesis ventured that non-intuitive uses of graphical elements may produce additional load for the learner. Some of the results of Chapter 6 showed that small specificity differences between systems were not mirrored by differences in learning outcomes in the predicted direction. The expectation for this study was to find non-intuitive uses of graphical systems in Euler system. This would help to explain the value of the specificity principle. The specificity idea combined with some measure of intuitive uses of graphical conventions would then be able to account for learner's relative benefit for the Venn, Euler and TARSKI'S WORLD systems.

The quantitative analysis demonstrates that learners do not have a very immediate or intuitive grasp of the representational conventions for each of the systems. The results show that the most specific system (Edinburgh Euler) had the most intuitive graphical conventions.



The only system that scored higher was the Euler\* system which replaces the shading in Euler with a star.

This result casts doubt again on the idea that specificity is a major component in deciding between alternative representations. The Venn diagram system is less specific than Euler but a better teaching tool. It also had less intuitive lexical components than the Venn system. Since we now know that the Euler system has more intuitive conventions than Venn we cannot use these results to explain the initial low results of the students learning from it.

The next study looked at individuals from the card sorting task and interviewed them with their card sorting choices available. The study illustrates learner's explanations for their choices in the card sorting task. The transcripts revealed a range of errors made by the twenty subjects. Many of these refer to problems interpreting the diagrams properly and others refer to logical errors interpreting the premises. The interview that follows is typical of the whole group and demonstrates almost all the errors found. The interview is atypical in that the subject was able to clearly explain their reasons for choice. Other subjects made similar errors but were less able to explain them clearly. This subject was one of the four graduate researchers in the sample who had not completed advanced mathematics study.

### *Interviews*

This next section reports selected interview material from an informative case study from the subject pool. The learner made incorrect assumptions that reflect the range of problems displayed by the group of twenty. These results enrich our understanding of the way in which poor learners approach these systems.

The methodology for these interviews follows in the tradition of approaches in the learning sciences, although the details of the elicitation method have been adapted to the purposes of this project. An interview approach is described by Gilbert, Watts and Osborne (1985) that they call "interview-about-instances". The method of the study reported here, draws from this and from other interview based approaches in the Phenomenography program (*e.g.*, Lybeck, Marton, Strömdahl & Tullberg, 1988; Lindström, 1980). The first of these projects provided most guidance to the methods used here. The Gilbert *et al.* (1985) methodology uses a card based elicitation process. They describe guidelines for developing new card sets and provide a template for the transcribed dialogue. Their approach is different from the one used here in several ways. The Gilbert *et al.* (1985) study investigates high school conceptions in physics. The cards illustrate situations that include non-criteria and criteria attributes of the concepts. The interview is structured around the sequence of cards presented to the student. The progression begins with clear illustrations of the concept, progresses to clear non-examples and ends with borderline examples and difficult instances.

The interview process uses open ended questions, non-evaluative responses and supplementary questions that continue to reveal the student's conceptions.

The approach used in this study began with cards based on representations of premises from five completely valid systems. The task of the student was not only to sort them into exemplar and non-exemplar illustrations but to sort them by order of preference for the particular premises. The ordering of choices was necessary to generate the quantitative data in shown in Table 27. The Gilbert *et al.* (1985) methodology does not appear to report the student's rating of 'prototypicality' for the illustrations. The purposes of their study were not to compare the systems for illustration, but to categorize the student's conceptions of force.

Our interview approach was also slightly different. The interviewer's objective in this study was to elicit justifications for the choices learners had already made. This is quite different from eliciting learner's understandings of the domain. The questions raised with subjects are intended to elicit justifications for their decisions. Most of the interviewer sentences begin with "So in that case..." or "So you think that..." and in a few examples, a question is asked to distinguish between two possible explanations for a choice. The approach could be described as rule elicitation.

The final contrast with the interview-about-instances approach lies in transcription methods. The method used here is very simple. The Gilbert *et al.* (1985) study includes tonal inflections, stressed syllables and duration of pauses. Dialogue in the following extracts are labeled by S indicating the subject and I the instructor. Within the dialogue, underlining is used to emphasise particular parts to the reader.

Apart from these several differences the study draws on the "interview-about-instances" approach. Both approaches are card based elicitation practices that emphasise learner preconceptions. Both are attempts to systematise a qualitative technique to improve reliability and both rely on card sorting as the primary means of access to student's models.

The progress of the interview followed a left to right top down traversal of the diagrams set out on a table top. Table 28 reproduces the order of choices made in this case study. The analysis of the interview transcripts, like the exploratory data analysis with the video data in the following study, takes several iterations to find the underlying themes. The sequential ordering is valuable since it preserves the preference order for individual conventions.

In the interviews referring to the card sorting task, we were trying to find out why the subjects had made the choices they did. More precisely, we wanted to know whether the learners were making correct and incorrect choices of diagrams based in their understanding of the diagrams, or on their understanding of the semantics of the logical statements they were asked to represent.

The preference for the star over the shading can be seen in the Table 28. Where a learner provided a correct premise diagram for a system, the name of system appears in the appropriate cell. A dash appears in the table when the diagram presented was not correct.



Table 28: Card sorting results for single subject in interview.

	Order of Choice			
	1st	2nd	3rd	4th
I (Some As are Bs)	V	-	-	-
E (No As are Bs)	-	-	E	E *
A (All As are Bs)	E	E *	-	-
O (Some As are not Bs)	-	-	V	-

Note: Dashed entries in the table represent invalid choices made by the subject.

The interview represents a case study of one individual who took part in the card sorting task. The subject showed several errors in the dialogue related both to misunderstandings about the logic and the meanings of lexical items from each of the systems. Five separate errors emerged in the interviews. Subjects using the Carroll system were often unable to distinguish the beginning and end of the regions referred to by labels (1). Learners using TARSKI’S WORLD had problems translating the terms from a given problem to the more concrete objects and predicates (2). More general behaviours include an order of preferred lexical items (3). Several learners failed to see the importance of shading (4). Intersection of circles to show common membership, and the star to show existence, are almost universally understood. Logical errors include extending the meaning of a premise beyond what it means strictly, for example ‘All As are Bs’ implies there are ‘No Bs that are not As’ (5). The following transcript excerpts introduce each of these errors.

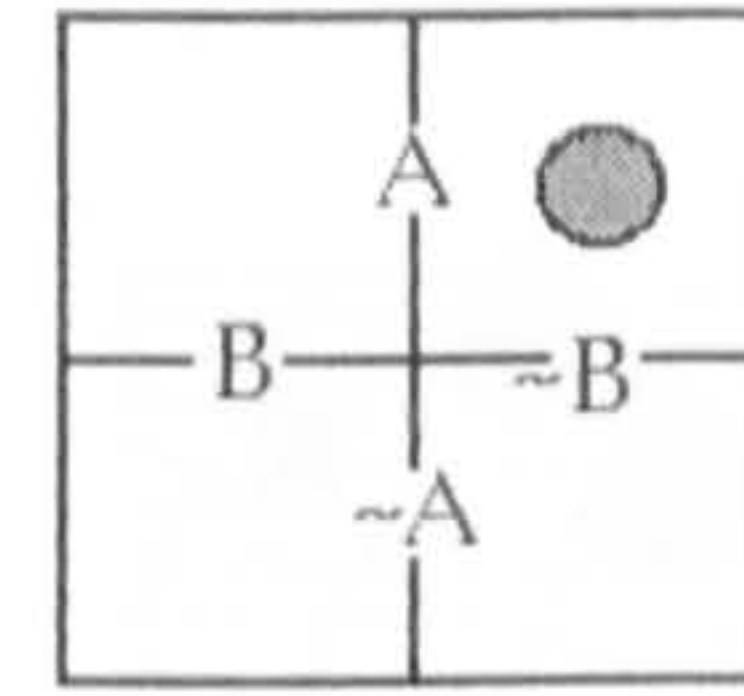
*Distinguishing the scope of the regions in Carroll*

Learner interpretations of the Carroll system often fail to see the proper scope of labeled regions. The label referent can be confusing because a single region may be described in one sentence as ‘A’ and then in the next as a ‘not B’. In reality, each region could properly be described by both terms. The region in the diagram in the dialogue that contains the shaded counter, would then be ‘A and not B’. Describing the region as part of the A rectangle and part of the ‘not B’ rectangle is a short-hand.



## Some As are Bs (1,5) second choice

- S ...because it is not clear to me where the division is between the As and the Bs, except for where you've got the letters, the actual areas...the actual spatial areas aren't clear to me.
- I So you said that you think it showed Some As are not Bs?
- S Yea.
- I And what was the sentence that was being attempted to be represented?
- S Some Architects are Builders... yea.. So really the better diagram would be where the circle is in up here (top-left) in the left hand quadrant.



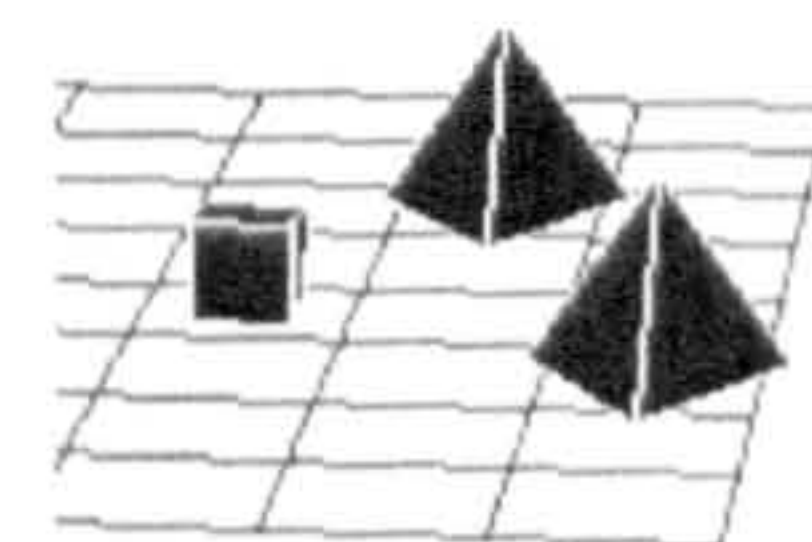
## Translating the terms to objects in TARSKI'S WORLD

Illustrations using the TARSKI'S WORLD format were rarely for any premise types (see Table 26). One reason for this is the difficulty subjects had with translating the objects and predicates in the representation back to the terms and predicates in the syllogism being solved. Both these two excerpts from an interview, which use the same picture, illustrate the problem learners have.

A fourth choice A-type premise representation using uses the TARSKI'S WORLD card. The learner's error involves mapping the A and B terms onto the predicates in the representation (A = Tetrahedron, B = Large). For All Architects are Builders the representation should show All Tetrahedrons are Large. In the software version this representation this could be achieved with: (a) an empty grid, (b) a grid with one large tetrahedron and nothing else or (c) a grid with many large tetrahedron and nothing else.

## All As are Bs (2) fourth choice

- S All these TW ones ever mean to me is that some of the things are represented by shapes and sizes. So if architects are represented by the larger, and builders are represented by the smaller.. they just share triangularity, so that means that some of them are part of... they share something in one set, one characteristic, but not in another.



a = tetrahedron,  
b = large

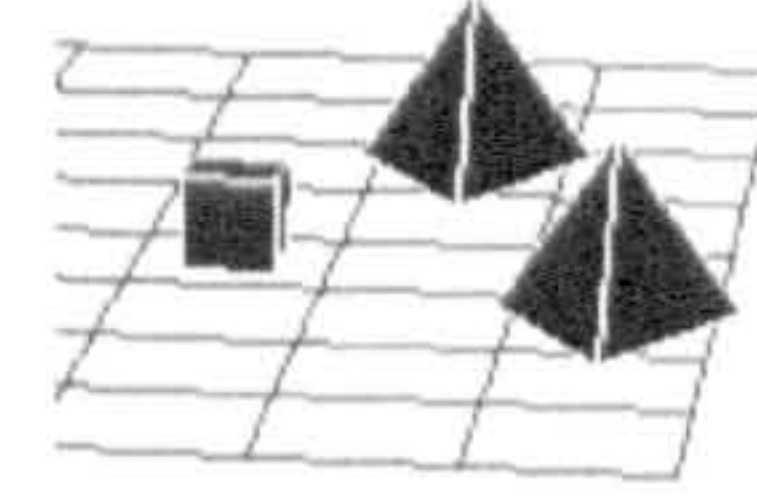


Some As are not Bs (2) fourth choice

*S Well...Uhhh... its confusing in a way because there are three things on there, there are two tetrahedrons and one cube. So architects could be the tetrahedrons, and the cube could be the builders and because they don't share either shape nor size... that's why I picked that one.*

*I Did you choose these pictures because you thought you should choose at least one from the tetrahedron group?*

*S If they were shared... if some architects were also builders then you would actually have shapes and sizes... you could have like a small triangle with two large, with two... or a small tetrahedron with two large tetrahedrons and that might indicate that there is some shared membership....(Sigh)...*



a = tetrahedron,  
b = large

The diagram above was chosen as the subject's final choice for the O-type premise and is one of the few instances of TARSKI'S WORLD being selected. The dialogue suggests there was little thought in the choice. Even though the A and B terms are clearly labeled on the cards as tetrahedrons and large objects, the subject fails to see the connection between the two levels of abstraction used in this representation. The representation is incidentally correct since there are no tetrahedrons that are large and so, No As are Bs. The user does not understand that the diagram is correct.

#### *The order of preferred items*

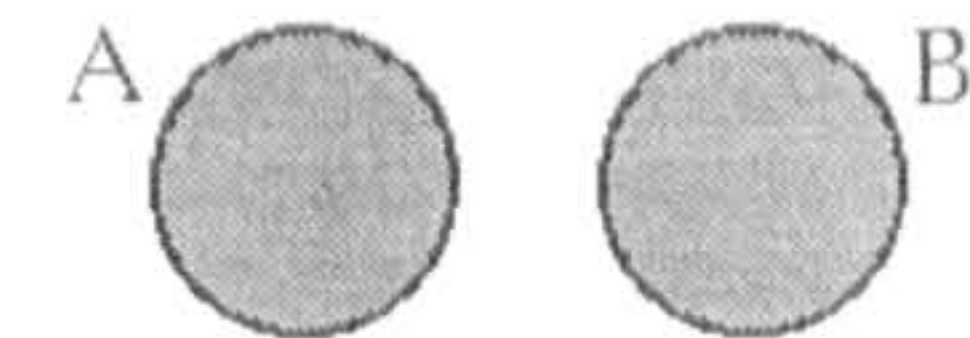
These extracts show learners making explicit statements that relate to their preferences for conventions. The prevailing preference (shown in Table 26) is for the star is liked more than the shading for showing existence. This is seen in comparing representations for Venn and Euler\* in the 'I' mood (which are the same diagram) with Euler which uses shading. This is also shown in the preference for Euler\* over Euler in the 'A' mood. In the following three extracts, separate circles (no intersection) is much preferred over shading to show no common membership between two premises.

No As are Bs (3) third choice

*S In this one you've got two exclusive groups, that don't interact. You've got architects over here and builders over here. And they both exist but they don't share anything.*

*I So they both exist, does that mean shading means existence again?*

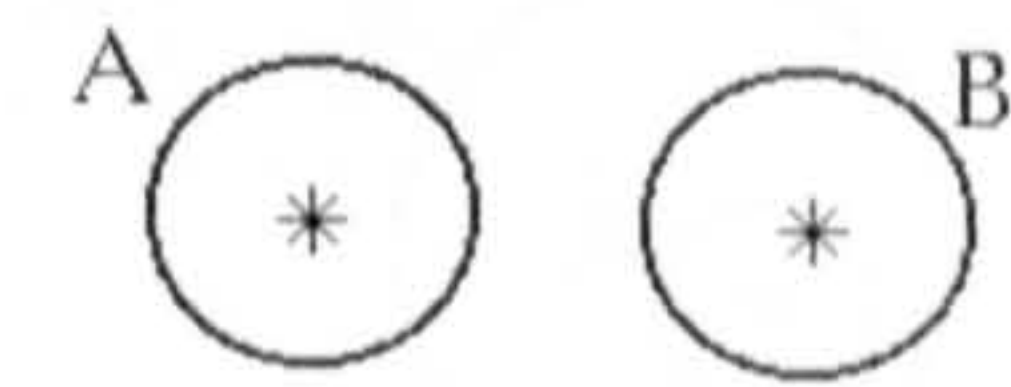
*S Yes, but they don't intersect, that's the most important thing.*





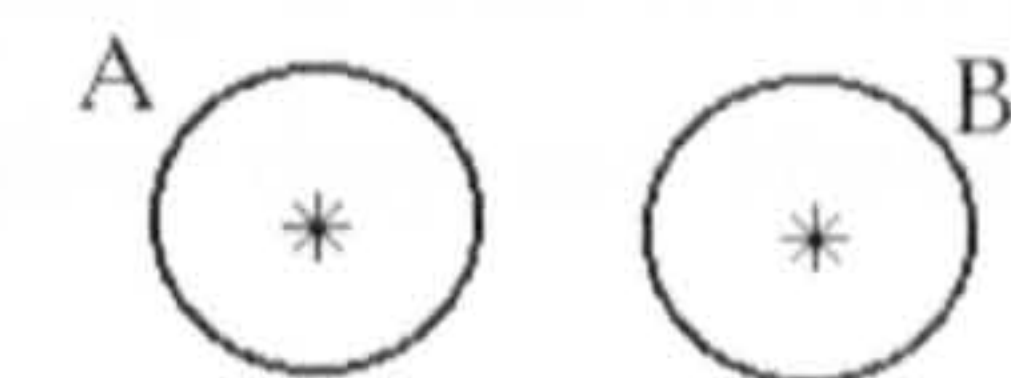
No As are Bs (3) fourth choice

- S* *It's the same with this, you've got the stars which indicate existence, but there is no overlap.*



Some As are not Bs (3) second choice

- S* *This one is the same again, the star is exactly the same as the shading...as it has been all along. The separated circles mean that A and B is exclusive.*  
*I* *Again there is no possibility for Some Architects being Builders.*  
*S* *That's right...(but) that's not what the picture is supposed to show is it?*

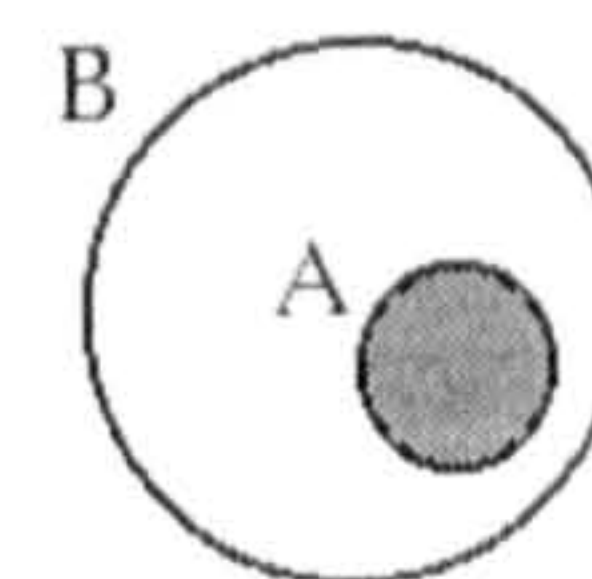


*Shading is often ignored*

The use of shading has provided problems for learners at several stages of this research. The preference that learners show for intersection (the previous description), is reinforced by the ignoring of shading seen in this excerpt. The high preference for intersection over shading (and even the star), is shown when these annotations are thought of as redundant. The shading, is so unimportant for this subject, that it would not matter to him if it was there or not. In earlier chapters we have called this bug, 'shading as redundant highlighting'

All As are Bs (3) first choice

- S* *... one circle is enclosed in the other, that means to me that all of the things that are architects are also builders, so everything that is in A is also in B. The shading actually in this case doesn't make very much difference to me, its just that one is actually inside of the other...*



This next example is quite explicit. The star is just an additional clue and the subject would have picked this diagram without any shading and without the star. He eventually accepts the benefit of the star but only as a redundant highlighting convention.



All As are Bs (3) second choice

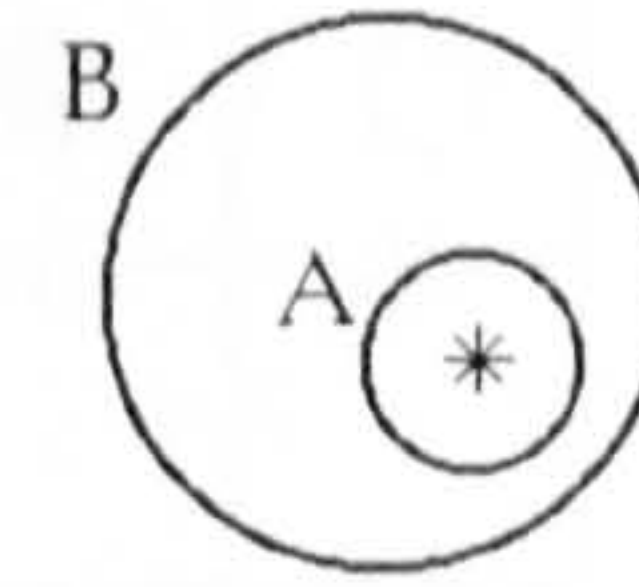
S *It's the same with this one. I would have picked this one... even if... its just the spatial thing, having the A actually physically inside the B means that all the things are parts of something else.*

I *So the star replaces the shading?*

S *Yea.*

I *So if the star wasn't there would it be better or worse?*

S *Maybe it would be worse, because I would think that there was nothing in either of the circles, that nothing actually existed in there. But that's just the additional clue.*



This description shows that the subject has an interesting view of what the areas which are not shaded mean. This is a direct consequence of the misunderstanding about shading. The subject simply believes that each region contains some members whether they are shaded or not.

Some As are Bs () first choice

S *If there are architects that are builders and builders that are architects then there are some that are both. That's why there is the shading in the middle.*

I *So what does the shading represent then?*

S *That there are people that are in that shading area*

I *So the shading represents existence of people.*

S *The not shaded parts are just architects and just builders, they are exclusive.*

I *So there are builders in the B circle...*

S *And there are architects in the A circle, but the ones that are architects and builders are in the shaded area.*

I *So the shading represents the shared labels of architects and builders.*

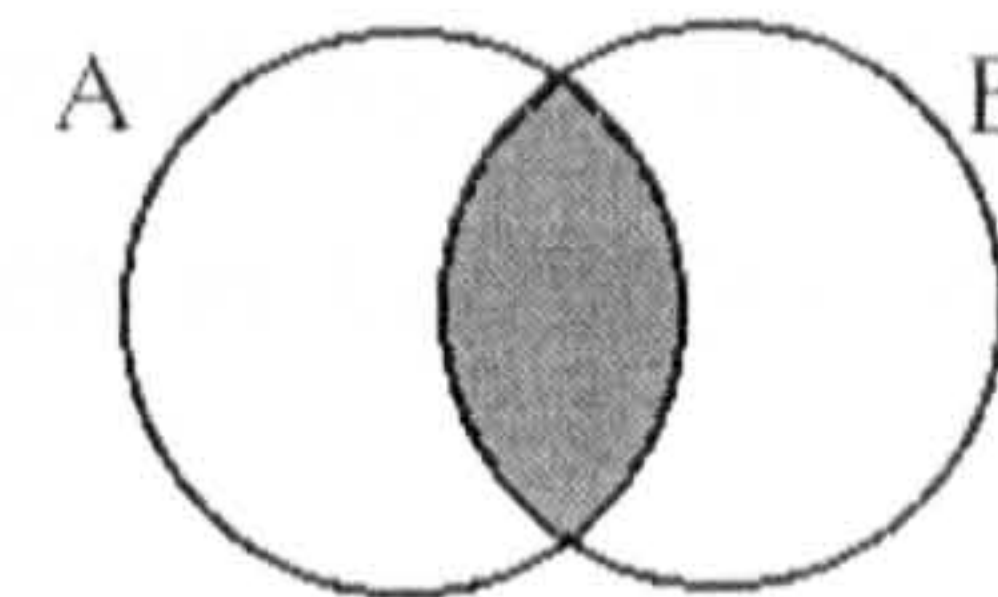
S *Yep.*

I *Do you think the intersection of the circles means anything? For example, if the circles were intersected but unshaded would that make any difference?*

S *Um...Yea.. but that would still be the same meaning.*

I *Would it be as good as the first diagram?*

S *No, this one is easier to understand with the shading.*

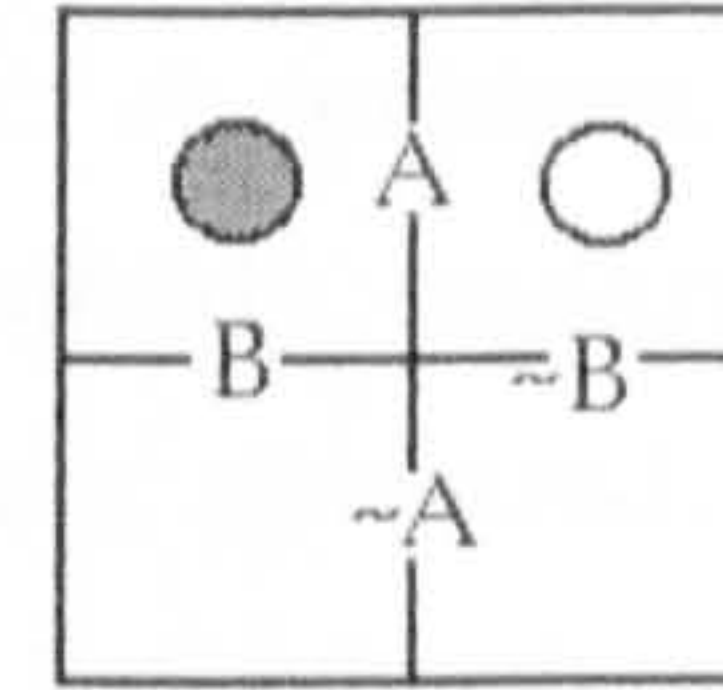


Another variation on the shading problem appears in the use of the Carroll system. The parallel of region shading in Venn, is counter shading Carroll. The suggestion that not shading an area might mean something different from shading an area is at last realised after raising the very explicit question in this transcript.



Some As are Bs (4) third choice

- I* So what does the shading mean for you here then?  
*S* The shading means that there are architects who are builders.  
*I* And what does the not shading mean?  
*S* Oh that doesn't make sense I shouldn't have picked that one.  
*I* So you think the shading and un-shading means something different now?  
*S* Yea, in this diagram because if I'm saying that the shading means that there is nothing there, um... then that would mean that there were no people in an empty circle ... and that is not true.. because there would be people who were architects and not builders .



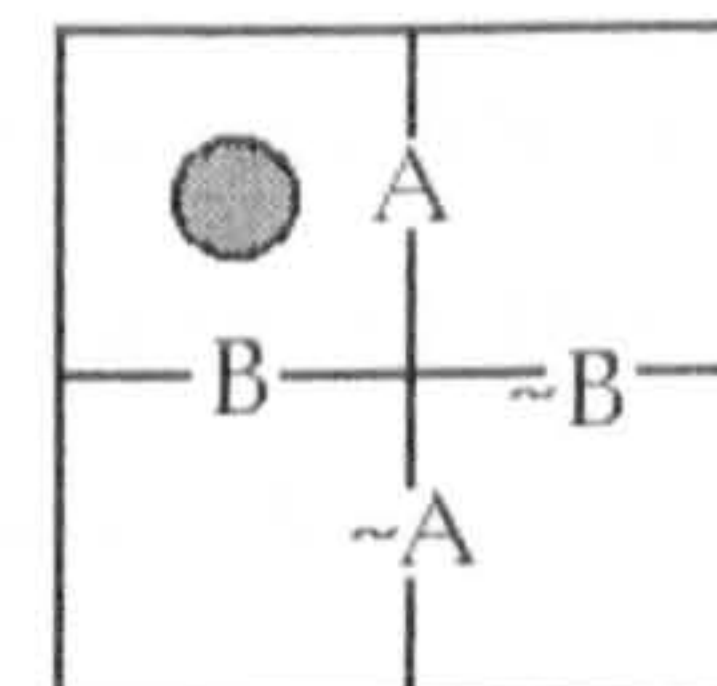
### Logical errors

Three logical errors were recorded in this interview. The first was a conversion-like error ( $\text{All As are Bs} \Rightarrow \text{All Bs are As}$ ), the second and third, problems with the 'Some' quantifier ( $\text{Some As} \Rightarrow \text{A few As}$ ) and ( $\text{Some As are Bs} \Rightarrow \text{Some As are not Bs}$ ).

The first example here was a third choice A-type premise. The error appears to be in the assumption that when All As are Bs, it follows that there are No Bs that are not As. This probably explains the missing empty circle in the upper right quadrant. This is indicated in the text by failure to recognise the implication of the last question. When All As are Bs then there cannot be any As that are not Bs and so an empty circle is required in the upper left region.

All As are Bs (5) third choice

- S* There is something that exists in this circle because it is shaded, that is both A and B. In this case the shading does make a difference because if this was empty then I would have said there was nothing in there.  
*I* Let me ask you a question, If Some of the Architects are Builders, and then you heard that some of the architects were not builders as well... So the way you figured it at the moment, how would you show that some of the architects are not builders?  
*S* There would be a shaded circle in the upper right quadrant.  
*I* So Some architects are not builders would be just a shaded circle there.  
*S* Yea.  
*I* So in that case you'd have Some of the architects are not builders and some of the architects are builders...  
*S* Yea.  
*I* If some of the architects are not builders, does that mean that all the architects could be builders?  
*S* No.  
*I* Can all the architects be builders if we don't know that some of the architects are not builders?  
*S* No... But that's just logic... It's only if you think through it that much...

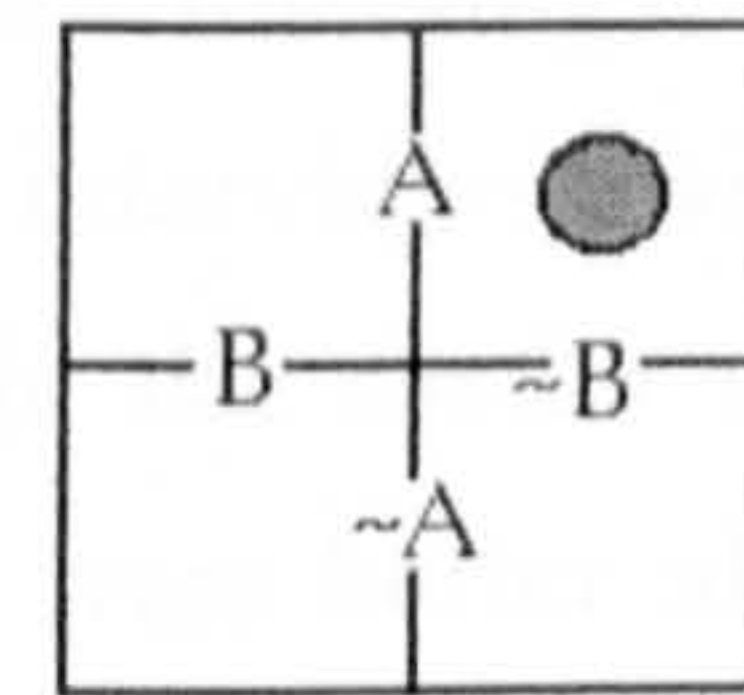




This subject's second choice diagram is from Carroll and is incorrect. The subject has the misunderstanding that a diagram (and presumably any statement), must show 'Some Bs are not As' as well as 'Some As are Bs' for it to properly show just 'Some As are Bs'. In the end the subject makes a better suggestion for the location of the counter. A problem with Carroll's representation seems to be that the scope of the labels are not properly understood which is to say that boundaries of regions named by each of the labels are confused. The extract is also an example of the first category of error (distinguishing the boundaries in Carroll).

Some As are Bs (1,5) second choice

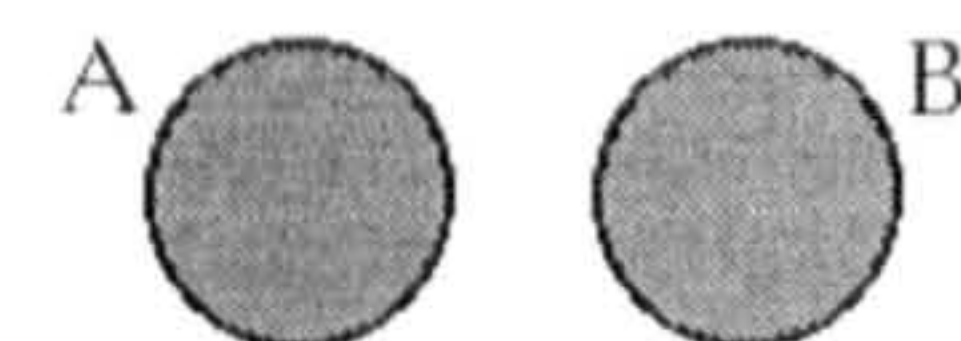
- S* (...) I don't really understand this diagram, but I'm assuming that what this means is..(...) I'm not sure...I know that symbol here is not—B ( $\neg B$ ), actually if this is the case then this circle here means that there are some architects that are not builders, and I think that is true of this diagram, but that is not everything that you want to say. So, it doesn't really show the best that some architects are...it would be better if that circle was right in the intersection of the four quadrants, because it is not clear to me where the division is between the As and the Bs. except for where you've got the letters, the actual areas...the actual spatial areas aren't clear to me.
- I* So you said that you think it showed Some As are not Bs?
- S* Yea.
- I* And what was the sentence that was being attempted to be represented?
- S* Some Architects are Builders... yea.. So really the better diagram would be where the circle is in up here in the left hand quadrant.



This next diagram is a first choice diagram for an O-type premise 'Some As are not Bs', and shows an over statement. It is true that when 'No As are Bs', that there is the valid implication that 'Some As are not Bs' however the statement on its own, 'Some As are not Bs' allows for the possibility that 'Some As are Bs' and does not commit to the statement 'No Bs are As' as this diagram does.

Some As are not Bs (5) first choice

- I* What about this one then for 'Some Architects are not builders'?
- S* The shading again shows that something exists, the As and Bs are separate and not members of the same group.

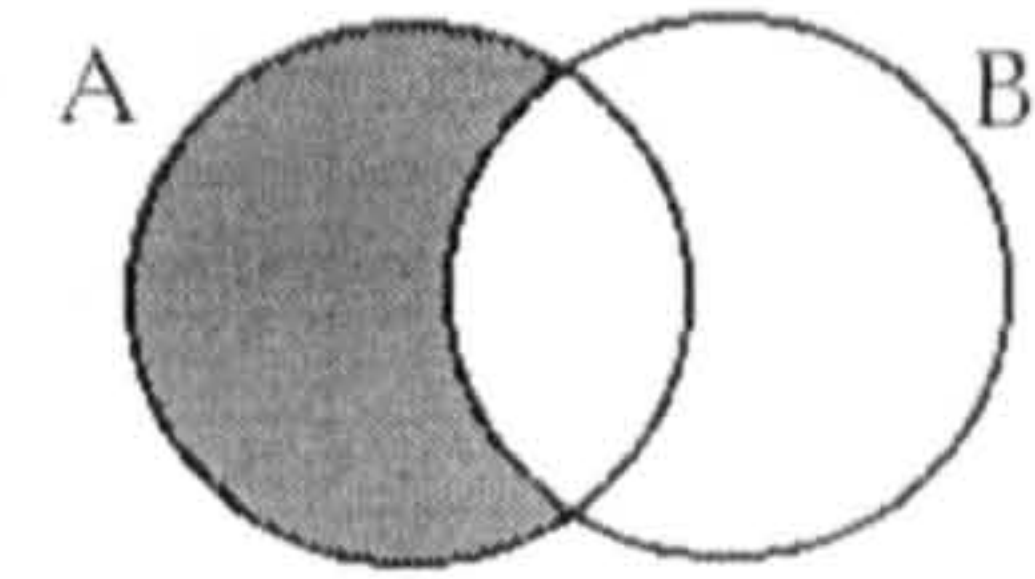


In this diagram the subject makes this choice last of all for this premise although it is the legitimate one for Euler and for the reasons that he explains. The preference is low because the subject prefers the overstated representation above.



Some As are not Bs (5) third choice

*S In this one the shading again shows that is where something exists. In this case though the representation is not so good because it might be that there was something in the intersecting region or the other part of B which is not shaded.*



Interviews revealed several faulty preconceptions of the diagram meanings. This particular transcript illustrates all of these errors better than others. The interviewed subject was one of the non-mathematically trained graduate researchers involved in the earlier study. The interview data appears to show the more mature non-mathematicians were better able to discuss their choices coherently in the interview situation than the secondary school non-mathematicians.

Errors reported from the interview data were counted and frequencies noted. This data provides a rough indication of the scope of each error. The frequencies give a rough indication of how many of the twenty subjects demonstrated the error.

An error with the Carroll system, shown by four of the twenty subjects, involved confusion over the scope of the regions used to locate the counters. This error is similar to the error in the Venn system known as the third term error. The problem in Venn manifests in omission of annotation from regions that include parts of the third term circle. The problem in Carroll is that learners fail to see each quadrant as the intersection of the column and row that adjoin it.

In Euler and Carroll another error involved interpreting the shading or the star always to mean existence of an individual. In Euler\* this is not a problem since shading does not appear in the system and the star indicates existence. Both shading and the star are used in the Venn system; shading to indicate nonexistence and the star to indicate existence. The prevalent errors occur in those premise types that use shading, namely A and O-type premises. In Euler this causes less error than in Venn since only shading is used to show existence. However the general problem with shading and recognising it's importance, probably contributes to the fact that Euler is less preferred than Euler\*.

Learners using TARSKI'S WORLD demonstrate a different set of problems. These errors occur while mapping predicates from pseudo-real objects to abstractions. TARSKI'S WORLD requires that premises are translated into orthogonal predicates such as Behind(a,b) and Large(a). This is necessary so that formally equivalent premises are not prohibited from being true, because of the nature of relations in the microworld. For example Large(a) is



inconsistent with Small(a) but not with Behind(a,c). This has the undesired effect of increasing the difficulty of the mapping from the premises and adds to learner's errors.

Results of the card sorting task show that subjects have a preference for the Euler systems and particularly the version of Euler that replaces the shading with a star. Results also show the Carroll and TARSKI'S WORLD systems are much less preferred. The card sorting technique does not currently include any way to measure significance between outcomes. Further development of the technique should provide this.

The card sorting task demonstrates that the representational primitives that make up the systems are subject to their own level of preferences. The subsequent interviewing yielded justifications for learner's choices of representation and produced two additional buggy categories. The first new bug describes shading as a redundant form of highlighting. The learner believes that the shading simply emphasizes what is already shown in the diagram. Without the shading the diagram would mean the same but might be less obvious. The second bug describes a logical error. This involves extending the meaning of the existential quantifier.

A third new finding is a variation on behaviour noted in the previous study reported in Chapter 6. Subjects in that study demonstrated a problem realizing the representations as embodying the *process* of problem solution. Users are unable to distinguish between a place holder in a diagram that constrains the possibility of an individual existing, and the actual representation of that individual in the diagram.

These additional insights to problems of interpreting the premise representations will be very useful in developing a more comprehensive model of errors and misconceptions in the domain. The result that is central to the theme of this chapter however, is most important. The difficulty in learning the lexicon for each of the systems was least for the unsuccessful system used for learning in the previous study.

The following section describes and reports a set of case training studies. Supporting learners with circle systems that can be directly manipulated can provide insightful interactions. The method of this next study begins to investigate learners' interaction with the dynamic parts of the circle systems.

#### *7.4 Training case studies & temporal analysis.*

Eight learners were selected from the pool of twenty subjects on the basis of their performance on the card-sorting task. The four highest scoring and four lowest scoring subjects were included in this study. The selection criteria do not strictly measure performance at syllogistic reasoning with diagrams. We believed however, that the range of abilities implicit in the selection would be adequate to cover a wide range of learners' activities with the diagrams.

The study involves the use of rich video data that is very time consuming to analyse. Ratios of sequence time to analysis time as high as 1:5000 have been recorded (Ritter & Larkin, 1994). Our strategy for learner selection reduces time needed for analysis but maintains good chances of useful data.

The training case studies do not compare learning outcomes from different conditions and there was no accurate timing to ensure equal treatment periods. Subjects were scheduled for periods of one and half hours but these sometimes continued longer.

Many time related issues are involved in interacting with a graphical diagram to learn some domain. These issues include: (a) the sequence of questions posed to the student, (b) the need for tutorial interventions by the instructor and (c) the repetition of teaching interventions in the sequence. Other more basic level processing activities also occur in activities with time critical aspects. These activities emerge more clearly in this study because of the video based data collection technique. These activities include, boundary tracing (Marr & Hildreth, 1980), overlaying and superimposition. The collection of the data in a video format allows for a depth in analysis of behaviour involved in learning from software. The continuous replaying and searching of video data allows for a cyclical development of theory, interpretation and evidence collection.

Such theory development can be done without the support of any specific software tools, but useful tools have emerged recently. One way to take advantage of the essentially graphical, linear and time sequenced data stored on video tape, is to look specifically at time based issues or to focus on time based aspects of research questions.

The exploratory sequential data analysis approach recently emerged (Baeker, Grudin, Buxton & Greenberg, 1995) as a way to deal with time based analysis. Exploratory sequential data analysis (or ESDA) is both a qualitative and a quantitative method (Sanderson, James & Seidler, 1989). It is described as any empirical undertaking seeking to analyse systems, environmental and or behavioural data -- usually recorded in some way -- in which the sequential integrity of events is preserved. The analysis of the data: (a) represents a quest for their meaning in relation to some research or design question, (b) is guided methodologically by one or more traditions of practice, and (c) is approached at least at the outset in an exploratory mode.

The previous studies do not adequately capture the dynamic aspects of circle intersection during the registration process. Each of the software designs supports contrasting approaches to learner's interaction with the graphics used. By storing learner's activities on video the study was designed to reveal which of these activities were helpful to the learner.

The dynamic semantics of the circle systems is especially relevant for the Euler system. In its current implementation the exploration of constraints between circles is the major form of activity for the student. While Venn does not explicitly need the same flexibility in



intersection of the circles it uses, it can still be demonstrated and manipulated by a user in this way.

These studies have shown that there is often a general confusion, bug or inability to separate the form and function of the individual circles. This phenomena is indicated by learner's diagrams with missing intersections between circles, creation of new unworkable conventions, externalizations that indicate the third term error and even indecipherable diagrams. This study is designed to find deeper evidence of the phenomena and to capture it in a format that can be replayed. The communication between the tutor and the student is collected in an audio stream that is also available as transcribed data and can be played along with video sequences.

There are many tools available for ESDA with video footage including: (a) SHAPA (Sanderson, James & Seidler, 1989), (b) Timelines (Harrison, 1995), (c) Videonoter (Roscelle & Goldman, 1991) and (d) EVA (Mackay, 1989). This study uses Timelines with a Macintosh computer and VCR with a 'Control-L' remote keyboard control. In common with each of the other video analysis tools, Timelines enables the tracking of events and activities with either instantaneous occurrence or with measurable duration. The tool will then calculate the number of repetitions of the same event and the total duration of an activity. Temporal relations such as precedence and overlap can be seen from a graphical representation of the data.

The analysis of video footage includes, events, intervals, actions and relations between actions. All these are applied to this particular field of teaching, learning and communicating with diagrams. Events occur or happen at points in discrete time.

$e \text{ H } t \leftrightarrow \text{event } e \text{ is Happening at time } t$

Two very important relations that hold for this study are those of precedence and overlap. Time intervals ( $T_1$  and  $T_2$ ), which are periods of time each from some  $t_{0-n}$  to some other  $t_{0-n}$  can precede each other if for every  $t_i$ , occurring during  $T_1$  and for every  $t_j$  occurring in interval  $T_2$ ,  $i$  always occurs before  $j$ .

$T_1 \text{ P } T_2 \leftrightarrow \forall i, j(((t_i \in T_1) \wedge (t_j \in T_2)) \supset (t_i < t_j))$

Similarly, an interval can be said to overlap another when there is at least one  $t$  that occurs during time interval  $T_1$  which also occurs during interval  $T_2$ . This can be expressed in the first order notation.

$T_1 \text{ O } T_2 \leftrightarrow \exists t((t \in T_1) \wedge (t \in T_2))$

These descriptions in the first order language are sufficient to cover events shown with Timelines software (Harrison, 1995) below.

The study used eight subjects taken from the pool of 20 from the card sorting experiment. The eight individuals selected from the pool represented the four furthest and four closest scores to the intended meanings for the graphical lexicon used in the systems. The strategy of selecting these extreme cases, minimizes the high cost of video analysis while maintaining the best chances for useful results. The four high and four low scoring individuals were randomly allocated to each of the other case conditions as illustrated in Table 29.

Table 29: The eight conditions in ESDA study.

Card Correlation	Condition			
	Euler		Venn	
	Supported	Unsupported	Supported	Unsupported
High	S1	S3	S5	S7
Low	S2	S4	S6	S8

Each subject participated in a one-to-one interaction with the instructor that was supported by either the Euler or Venn representational systems. Half of the subjects used the support of software versions of these systems and the others used paper and pencil versions. In each case a tutor provided direct support and discussion with the learner during the process. The tutor intervenes when mistakes are made by the learner and when learners reach a halt in their attempts to solve a part of the problem. The tutor often provides hints and asks leading questions intended to direct the learner towards successful use of the tools.

Sometimes the learner and the instructor engaged in communication that included sketching circles at the same time as describing the solution or explaining what was different about their diagram and the tutor’s diagram. Communication with gestures often occurs with no marking on paper. A similar activity occurs when learners use the software. The equivalent of moving a pen around, to point at parts of a diagram on paper, is to use the mouse and cursor with the software.

The video data was collected with a standard full size VCR camera onto 90 minute VHS cassettes. The Timelines software is able to read the timestamp from the video tape, to locate and relocate the position of an event and to produce graphical Gaant charts from the data.



7.4.1 Results

Results of the case study training sessions are descriptive. The behaviour of each learner is compared based on their selection criteria and the treatment they received. The study was designed to capture data in a rich format that requires time consuming analysis. The methodology provides a way to form detailed and relatively intimate pictures of the individual learners and their use of the tools.

The descriptive statistics shown in Table 30 include total duration of activities, their frequencies and learners’ problem solving speed. There are thirteen listed activities and events including bugs, useful and unhelpful behaviour. This data also includes key presses collected to show some activities in the software supported cases.

The timelines software generates graphical output similar to Gaant charts. A modified version of this output is shown in Figures 21 & 22 below. The graphical output from the timelines software includes two versions of the same data; one of these is removed from the figures presented below. Additional annotations are made to the diagrams shown. The problem number has been added to the first row between the start problem events. The number of the subject is added to the illustrations in the lower left corner. Poor card sorters and good card sorters are those who scored low and high in the card sorting task.

The categories of events and activities represented in the Timelines software are not all possible in each of the systems. Table 30 shows impossible activities as grayed out cells in the table. The complete range of activities tracked in the study include:

- SP*    **Start Problem.** This event is initiated each time a new problem (syllogism), is begun.
- Sh*    **Shading.** When the subject shades an area of a diagram, this activity is recorded.
- TO*    **Term Order Error.** When a subject demonstrates behaviour indicating this error, the activity is recorded. The conditions for this behaviour are listed in, Chapter 6, Table 21.
- TT*    **Third Term.** The conditions for this error are shown in, Chapter 6, Table 18.
- ES*    **Empty Set.** The conditions for this error are shown in, Chapter 6, Table 20.
- \*M*    **\* Movement.** The conditions for this error (which only occurs in the Venn system), are shown in Chapter 6, Table 19.
- P*      **Pathological.** Diagrams that are not interpretable by any bug diagnosis tools.
- GS*    **Gestural Sketching.** This activity refers to continuous movement of the pointer, finger or pencil. The speaker indicates areas of the diagram that are important. The activity may occur in communication with others, or alone. The term ‘diectic sketching’ was used after (Logan, 1995), however, ‘gestural’ is a more inclusive term and has been adopted.
- I*      **Intervention.** The instructor steps in and the activity is recorded.
- EC*    **Explore Constraints.** In EULER, there are limits on the movement of the circles within the diagram. When the student is seen to explore these limits, the activity is recorded.
- MI*    **Missing Intersection.** A subject creates a diagram with a missing intersection in the Euler system.

- Cb** *Combine button.* A button in the EULER system, which registers the two premise diagrams into a minimally intersected diagram.
- Cf** In EULER, the button labeled 'Is it right' allows the user to confirm that the diagram created has the maximum number of intersected regions. Whenever this button is pressed, the event is recorded.

This complete list of behaviours is a combination of the activities and events recorded from all learners including, good and poor learners, as well as users of Venn and Euler, with and without software. The graphical output in Figures 21 and 22 show these behaviour categories in the column to the left of the Gaant output. These activities are the user created labels of active buttons in the Timelines software. Iterative study and theory development allow a detailed analysis of the data. New buttons can be created during the analysis and the data can be reviewed with the new buttons. Clicking on these buttons records the beginning and end of activities. The graphical output is produced when the software calculates and draws graphical bars for the beginning and end of each activity or event. A visual impression of the relationships between each of the bars helps to see patterns in the video footage that might not be obvious from the raw data. The graphical output is a communication device as well as an aid to analysis.

The individual behaviour of one learner can illustrate their problem solving activities through time. The image provides a visual characterization of each learner. For example, student 4 in this study (shown in the following figure), was drawn from the card sorting pool as one of the four poorest card sorters. He used paper and pencil with the instructor to learn the correct method of Euler circle use. He started five problems and produced three pathological diagrams in the first two questions. Pathological diagrams are impossible to decipher (see Chapter 6). An intervention by the tutor immediately follows the learner creating these diagrams. After correcting the first diagram the learner makes a shading error. This is the first shading error of nine made in all. The numerical data produced by Timelines summed to a total duration for all his shading errors of 510 seconds. The average duration of shading errors shown by this learner lasted 57 seconds. This is measured from the moment it was first indicated until the moment it was corrected or another activity begun. The graphical output shows that in the fourth problem the learner begins to communicate with the tutor referring to the diagram through pointing and tracing boundaries. This behaviour is unusual for the two learners using Euler with poor card sorting scores.

There is no space or reason here, to characterize all the learners explicitly in this way. The diagrams are able to illustrate this information more concisely than description can. The diagrams illustrate the patterns of comparison between each of the treatment cases.



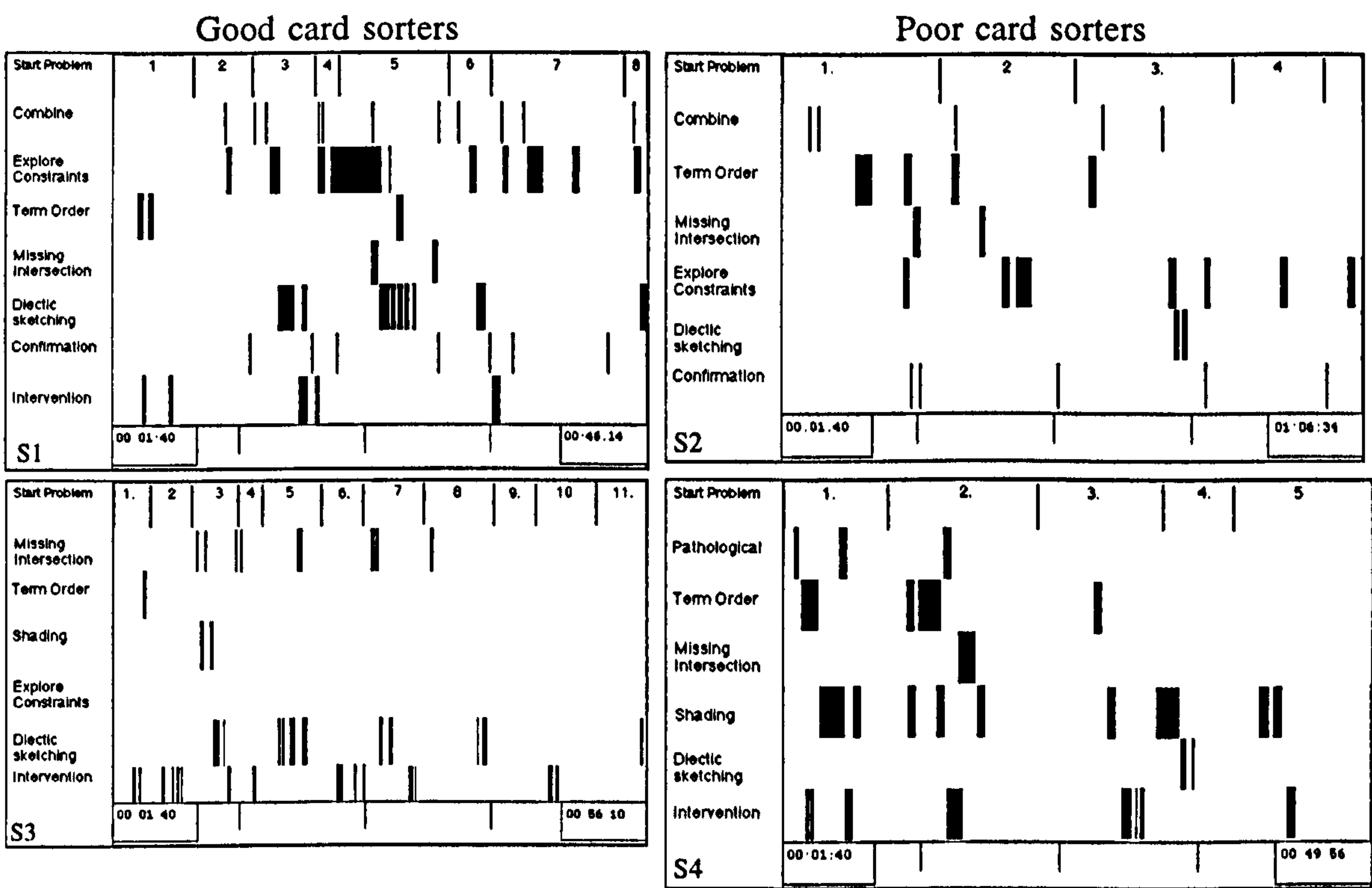


Figure 21: Timelines data from Euler conditions, (S1 & S2 are supported with the software, S3 & S8, used paper and pencil).

The graphical representation of the data is particularly good at illustrating patterns in the data. The illustrations show different densities of interaction between the learner and the instructor. The number of annotations on the chart illustrates the overall quantity of interaction. The good card sorters in Figure 21 are engaged in more activity (mostly interaction), than the poor card sorters. This ‘density’ of activity emerges as a pattern during comparison between the left and right charts.

The duration of activities in the good card sorters activities also appear to be shorter although this is misleading. The lengths of lines representing duration of activities are much shorter than those on the right side. The total duration for all these conditions is about the same, so when more problems are dealt with, the interaction cycles are often faster. Good card sorters complete the problem solving process more quickly but show similar patterns of interactions.

Comparisons between the upper and lower charts in Figure 21 show that more problems are attempted with the paper based treatments. The software slowed the process of problem solving.

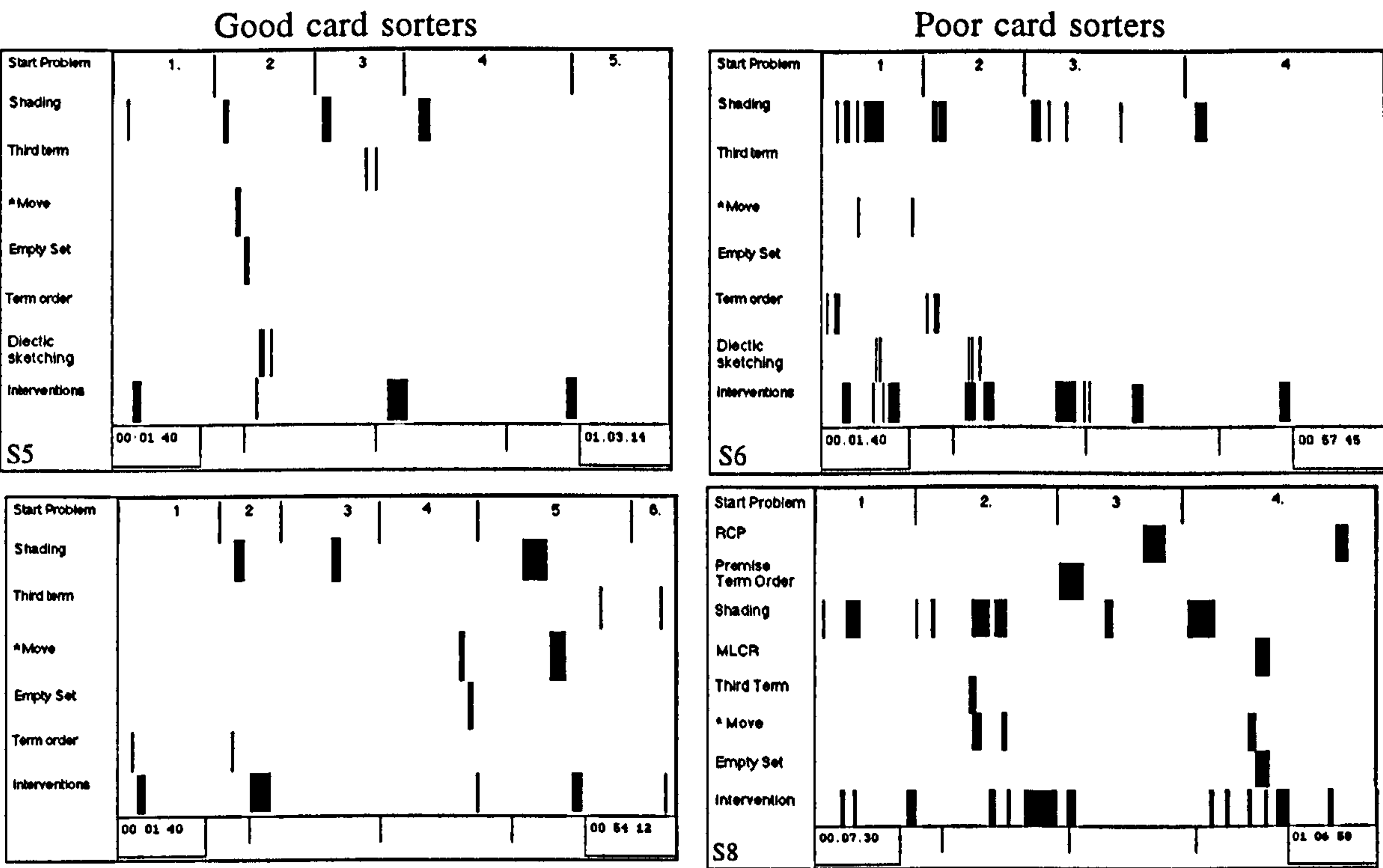


Figure 22: Timelines data from Venn conditions, (S5 & S6 are supported with the software, S7 & S8 used paper and pencil).

Bugs found and tracked with the Timelines software with users of the Euler software included: (a) neglecting to shade any of the diagram, (b) neglecting to display any of the premise diagrams, (c) neglecting to label circles, (d) incorrectly interpreting the term order from premises, (e) omitting possible circle intersections, (f) making diagrams with no resemblance to the required diagram, (g) inventing novel conventions that sometimes worked and (h) invented conventions which were not useful. In the Venn group seven bugs were noted and frequencies calculated: (a) the third term error, (b) the star movement error, (c) the empty set error, (d) the premise term order error, (e) the pathological diagram, (f) the spurious conventions, and (g) representing the conclusion as a premise.

The shaded cells in Table 30 below indicate errors in each column that are not possible for the system indicated by the row. It is not possible, for example, to incorrectly shade regions with the Euler software. The software does all the shading for the user. The absence of these impossible errors does not indicate better performance by those learners using the software. Removing the option to make those errors can be thought of as a particular form of instructional strategy or scaffolding of the learning process. However, learners using the software cannot make these errors, so have no opportunity to benefit from either tutorial intervention or from feedback when their conclusions are mistaken. The learners without software support made use of tutorial interventions slightly more than those learners who used



the software. Learners using the software averaged 5.2 episodes of intervention compared with 10.5 episodes used by learners without the software.

Some interesting differences emerged between the good card sorters (S1, S3, S5 & S7), and poor card-sorters (S2, S4, S6 & S8). Good card sorters generally started more problems than poor sorters. Average number of problems started by good card sorters was 7.5 compared with 4.5 problems started by poor card sorters. The number of interventions made by the instructor is an indication of learners' need for help. The results do not clearly show that poor card sorters needed more help. However this is partly because one learner (S2) refused the support of the instructor. Without this learner the difference in interventions between the poor card sorters and the good card sorters is more clear. Poor card sorters required an average of 11.3 interventions during teaching and learning, while good card sorters required only 7.5 interventions.

Both these results indicate that outcomes of the card-sorting task were representative of users' confidence with the systems. Although the study did not track changes in performance, the transcripts also indicate this group was more successful at correctly solving the problems. The results show the card sorting task was representative of learners eventual success at solving the syllogisms.

The practice of gestural sketching was most prominent in the Euler conditions. Twenty-six separate episodes of gestural sketching occurred among all four Euler cases. These episodes averaged around thirty seconds at a time, indicating an important learning strategy. Gestural sketching appears to be a good strategy for helping to internalize the abstract model of a system. The practice helps the learner to use the particular convention of intersection of closed curves to best advantage by helping to focus on the 'continuity' of the circle boundaries. Learners in the Venn group showed far fewer examples of this behaviour. A total of only seven episodes that lasted around twenty seconds each. Circle boundaries may be clearer in the Venn system than in the Euler system. The static method of showing Venn with the three circles permanently fixed in one position may be less confusing to learners. Clearer use of circle boundaries may lessen the need for gestural sketching for users of the Venn system.

Shading errors appeared in both systems although not in the software supported Euler system since the software prohibits this. More shading errors occur with less able learners in both conditions particularly when considered as a proportion of the number of problems met.

These bugs are the beginning of a learning hierarchy, extendible to a more detailed cognitive model of learning to perform syllogistic reasoning by the internalization of external graphical systems.

Performing a reasoning task with a graphical representation requires a number of specific tasks. Some of these are not specific to reasoning with diagrams, but are common to general problem solving: selection, matching, sorting and evaluating states from schema.

Generating new states from known constraints in the graphical representations requires special graphical problem solving skills. The procedures for generating new representations involve transformations of the image. These transformations must maintain the constraints of legitimate antecedent and consequent diagrams.

Other more visual procedures include edge, arc and boundary tracing, scanning and annotation in the construction and evaluation of an image. A model of problem solving with interactive graphical representations also needs a higher level monitoring and control process to switch appropriately between each of the processes of construction, scanning, evaluation, and making transformations. The diagram shown in Figure 23 models the general space of errors made while learning to perform syllogistic reasoning with the diagrammatic systems.

Table 30: Errors in 8 Condition study.

		Categories of Behaviour					
		SP	Sh	TO	TT	ES	*M
Euler-SH	S1	8	-	3(145)	-	-	-
	SL S2	5	-	4(218)	-	-	-
	UH S3	11	3(35)	1(28)	-	-	-
	UL S4	5	9(510)	4(265)	-	-	-
Venn-SH	S5	5	4(175)	0	2(35)	1	1(54)
	SL S6	4	11(225)	4(85)	0	0	2(26)
	UH S7	6	3(219)	2(23)	2(25)	1	2(256)
	UL S8	4	8(725)	1(156)	1(53)	1	3(115)

		P	GS	I	EC	MI	Cb	Cf
Euler-SH	S1	0	9(386)	5(154)	10(630)	2(88)	11	7
	SL S2	0	2(89)	-	7(240)	2(63)	5	5
	UH S3	0	13(365)	16(145)	0	8(123)	-	-
	UL S4	3(211)	2(45)	8(240)	0	1(105)	-	-
Venn-SH	S5	0	20	4(171)	-	-	-	-
	SL S6	0	5(56)	12(354)	-	-	-	-
	UH S7	0	na	5(183)	-	-	-	-
	UL S8	0	na	13(418)	-	-	-	-

Note: SP = Start Problem, Sh = Shading, TO = Term Order, TT = Third Term, ES = Empty Set, \*M = \* Movement, P = Pathological, GS = Gestural Sketching, I = Intervention, EC = Explore Constraints, MI = Missing Intersection, Cb = Combine button, Cf = Confirm Button. Numbers in parenthesis are total durations for period categories (in seconds). Numbers outside brackets are occurrences of events. Shaded areas represent errors which are apriori impossible.

Errors in this map are grouped under the main headings of, (a) diagramming errors, (b) language errors and (c) belief errors. The studies reported in the thesis have mostly found diagrammatic, language and belief errors.

The project has paid little attention to diagram processing errors. The work scratchings made by learners may not be the best place to look for problems of this kind. Interviewing learners after the card sorting task allowed learners to explain their misconceptions of



diagram meanings. Processing errors are more intangible. Learners may find it difficult to assess the limits of intersecting regions. The strong circle shape influences the perception that the Venn diagram consists of three intersecting circles. The circles could be interpreted perversely as seven three sided figures with common edges. Although learners *recognise* the Venn diagram as three intersecting circles, they are not always able to treat the diagram as three circles. The third term error may in part be due to this problem. Learners who make diagrams like those with the third term error, may actually be making errors in processing the diagram.

The biggest contribution of this chapter may have been the description of activities that support diagram processing as well as the description of errors made in diagram processing. Further development of methods to study learners' use of interactive diagrams will bring better design guidance to improve both the diagrams and learners use of them. Studies designed to provide these results are planned for future work. Detailed analysis using video data collection and possibly eye tracking software may reveal differences in learners' attention that parallel their successful hand-eye activities like gestural sketching.

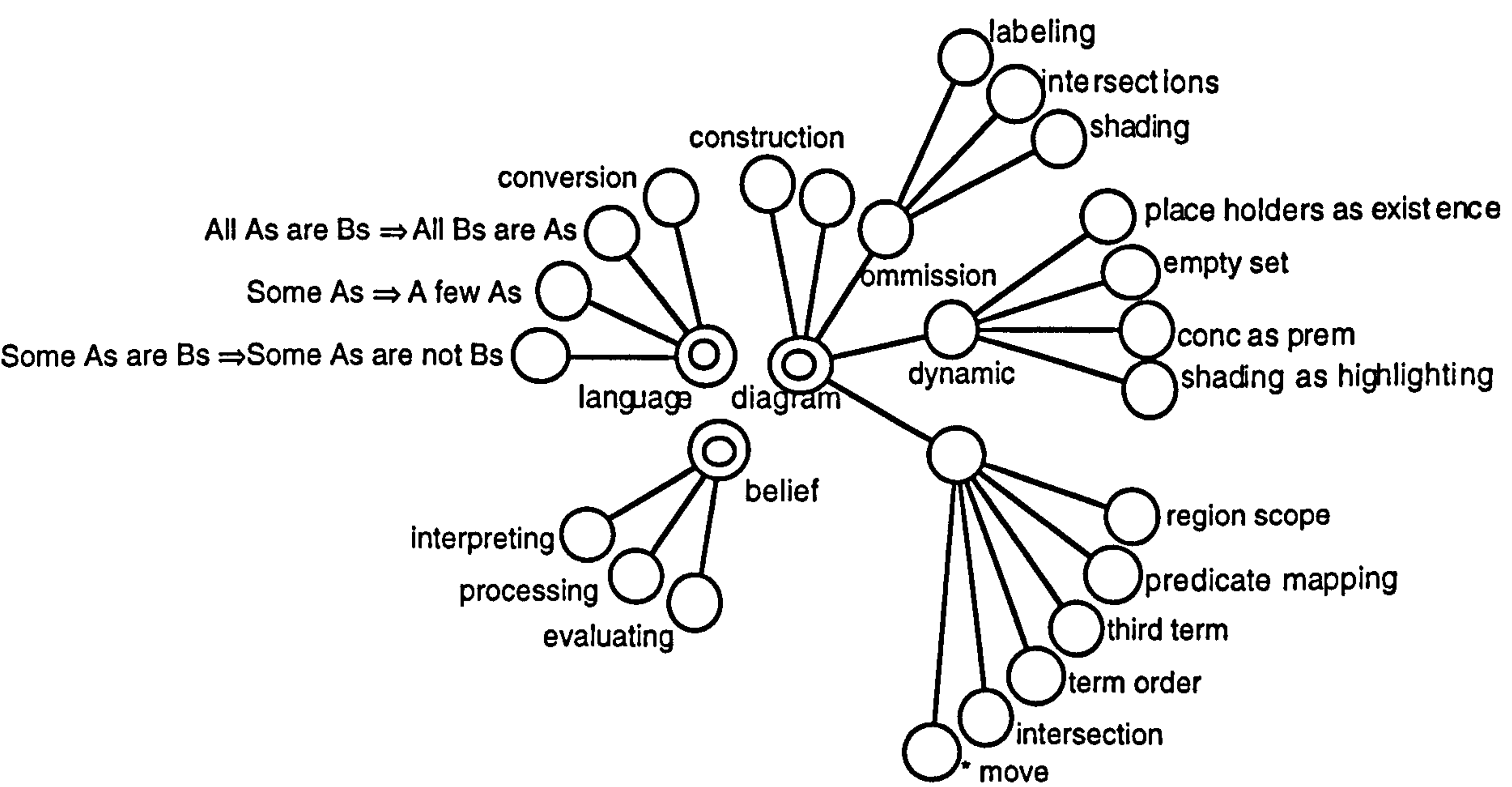


Figure 23: Map of student errors in syllogistic reasoning.

The act of pointing or gesturing was a very noticeable activity among six of the case studies. Learners using the EULER system showed more gestural behaviour than those in the Venn conditions. However the largest difference in gestural use was between the high and low correlation card sorters. Good card sorters did up to nine times more gesturing, pointing and

following boundaries of the circles. These good card sorters showed beneficial and proficient behaviour in several ways (see above). There is very good indication that the habit of following boundaries is a beneficial activity for integrating and learning the domain.

Investigation of gestural and mixed gestural and voice input is a new research area in multi-modal interface design. Some early findings in this field have already shown some benefit of gesture recognition systems in immersive learning environments (Ganschow *et al.*, 1994; Loftin, Engelberg & Benedetti, 1993; Reigian & Shebilske, 1992). Use of the term gestural sketching stems from work by Hauptmann and McAvinny (1993).

The human visual system is primarily a boundary or edge detection system used for isolating objects from noisy scenes. The use of gestural communication has been divided into gesticulation and virtual object manipulation by Hauptmann and McAvinney (1993). They speak of gesticulation as a way of adding emphasis to spoken communication. Virtual object manipulation describes virtual operations without a typical direct manipulation input device such as a mouse, using for example the 'data glove'.

A third form of gestural communication describes the use of gesture to depict graphical objects. We have called this "gestural sketching," and believe that the activity is useful to learners working with diagrams.

Since the visual system discriminates between objects by making inferences based on shading and intersection, the use of diagramming systems that use intersection to show complex relationships may need additional support. Learners who turn out to be good at using the graphical systems, often trace around the boundaries of the circles with fingers, pointers and pencils. This practice may well help to direct the learner to the intersections of the circles, and hinder processes normally used to look for objects in noisy images.

Using the circle systems for communicating the relations between terms in a premise involves hypothetical constraints in the diagrams. Diagrams may be manipulated while maintaining the truth of the premises. All the systems used in the studies have these constraints. The Euler system, however, makes exploring those constraints the primary learning experience for the user. It is difficult to show these kinds of dynamic constraints on paper. This kind of interaction is unique to computer based graphical displays. Continuously manipulable abstract objects that display constraints relevant to the important underlying semantics are not possible in any other mode of communication

The exploratory sequential data analysis contributes to an understanding of the support that graphical systems of representation can provide for the learner. The data particularly shows the benefit of constraint exploration and the benefit of systems that support constraint exploration.

Results of the study indicate that a combination of better guidance with the exploration of constraints would benefit learners. Progressively removing the guidance by allowing the



learner to create their diagrams without the support of built in conventions would allow the learner the opportunity to make mistakes needed for learning.

The results illustrate useful externalization techniques used by successful learners both for communication of the domain between learner and tutor, but also in reflective problem solving. The act of gestural sketching appears to support aspects of image processing like properly following the boundaries of circles.

The Euler system appears to be more successful in this study than the study reported in Chapter 6. This may be for several reasons. Teaching materials produced in the first study were designed for each of the software and diagramming conditions. Venn diagrams illustrate examples to learners who later used the Venn software and Euler diagrams for those in the Euler condition. The case training studies in this chapter involve the tutor in intensive conversations with learners using diagrams. The card sorting study shows that the conventions used in the Euler circle system are closer to what subjects expect them to mean than for the Venn system. This study did not take into account any quantitative measure of success in the learners improvement in skill level so we do not know if the Euler system really proved to be more useful than Venn. The study does show that less errors were shown with Euler than with Venn. The study appears to indicate that with the right kind of tutorial support, the Euler circle system can be more supportive than Venn. This is clearly at odds with the results of Chapter 6 but not with the predictions of the specificity principle.

### *7.5 Discussion.*

A series of three studies dealt with two surprising phenomena that emerged from the study reported in Chapter 6. The study in Chapter 6 showed that learners using the three software applications, TARSKI'S WORLD, EULER and VENN achieved remarkably different improvements in reasoning skills. The first surprising difference showed learners studying high level Mathematics in secondary school did much better than similar learners studying humanities subjects like Geography and History. The performance of non-mathematicians actually declined apparently as a result of using the software.

The first study reported in this chapter administered a test of cognitive style with ten mathematicians and ten humanities subjects. The hypothesis claimed that humanities learners made use of semantic elements in the problems to help reach conclusions. Such a strategy is useful in real world reasoning. The diagram systems quickly show that relying on beliefs about the empirical truth of statements to solve syllogisms with empirically untrue premises, results in invalid conclusions and poor performance. If these learners rejected their use of semantic cues and had not yet learned to replace it with a better strategy, this could explain the deterioration in performance shown by non-mathematics trained learners.

The study investigated the use of semantic cues by learners with non-mathematical training. The spy-ring test for cognitive style provides data to help inform this issue. The test found non-mathematics learners correlated with a 'holist' style of learning. Definitions of the holist style describe learners making use of semantic cues. Furthermore, the mathematics trained learners also showed strong correlation with the alternative 'serialist' style. These learners also scored highly in the 'versatile' style that allows learners to move freely between approaches that might be better in different situations.

These results do not completely confirm the hypothesis, but show a difference between groups that is consistent with the claim. The differences in style do not inevitably lead to rejection of learning strategies. The hypothesis could be more definitively tested by providing learners with problems consistent with empirically true premises but with increasing difficulty. If non-mathematics trained learners continue to succeed at these problems this would help show semantic cues contribute to performance more in this population.

The second surprising finding from the study in Chapter 6 was the mixed results of predictions made by the specificity principle. Part of the results confirmed these predictions. Learners using the highly expressive TARSKI'S WORLD software showed little improvement. However comparisons between the more similar circle systems appeared to contradict the specificity principle.

The second study reported in this chapter was based on the hypothesis that this difference was due to differences in difficulty for learners understanding parts of the graphical lexicon used by those systems. The study set out to discover which system had the more difficult lexicon. A unique form of card-sorting captured learners' preferences between the competing diagrams. The result of this study showed strong differences in preferences between the lexicons. However, the hypothesis was not confirmed because learners' preconceptions of the diagrams were closer to Euler's, the less successful system in Chapter 6. Each of the subjects was interviewed about their choices. This process revealed several mistaken preconceptions about the diagrams that could be incorporated into better teaching.

The result of the card sorting study left no explanation for the failure of the Euler system in the study reported in Chapter 6. The following study used a very detailed and data rich method to study learners using the software with an instructor. Learners were video taped while using each of the software applications and the video later analyzed with the Timelines software.

The training case studies provide a detailed look at some time based differences between learners using the systems. Subjects who did best at the card sorting, also answered more questions, needed less tutor support and showed fewer errors than the poorer card sorters. Learners without the software attempted more problems than those with software support. This study led to a model of learning syllogistic reasoning with diagrams, expressed as a concept map. Although the study design did not track differences in performance



changes with the systems, learners using the EULER system appeared to benefit from exploring the circle constraints. The training case study does not clearly show that the EULER system was more successful than VENN. However it does show that with good tutorial support learners showed satisfaction from the use of EULER. It is possible that the results of Chapter 6 were partly due to poor instructional materials.

The final chapter discusses the contributions of the project and the audiences that may find the results useful. The methods used in the project and the results may have been stronger in some areas. These limitations are discussed. The chapter outlines several opportunities for further work that could be done in this area.

## 8 Discussion & Conclusions

This chapter draws together the successes and contributions of the project and discusses some of its limitations. Several research questions arose during the project implementation. Shortage of time and relevance to the original research question, helped prioritize the study agenda. A description of several further projects includes work that emerged directly from the thesis results and work not possible during the project.

The contributions of the project are organised by field or discipline, and by the scale of their importance. Researchers in several different disciplines are likely to benefit from each of the project outcomes.

Some future work projects are directly related to the thesis results and others are more indirectly related. Some small experiments are planned that might require a few months of preparation and analysis and others are much larger questions that could grow into multi-year dissertation projects.

The exploratory methods used here have led to findings that must to be applied to new technology based learning systems to meet their full potential. The catalogues of bugs, misconceptions and errors in learning reasoning skills are ready to be built into new learning systems. The full benefit of these findings will be determined by the benefit provided to the learner using these systems.

Successful learning with diagram based technologies involves a complex interaction of several factors. The expectations, beliefs and approach of the learner all contribute to their success with the tools. Graphical systems can be designed with the learner in mind and there are certain guidelines that support good design. The project has explained and tested one guideline designed to improve the tractability of system semantics. There are other factors involved in the practical task of choosing a notation to teach with.

'System extensibility' is one such factor. Learning the lexicon of a representation almost always requires an investment of considerable effort. Some systems can be easily extended to meet the needs of increasing sophistication in the domain. Other systems may quickly become cumbersome or unnatural to use. All the diagrams used in this project, represent a small part of logic, and soon become unusable when other fragments of logic are added to a curriculum. These representations must be replaced with notations more appropriate to the more complex material. The new effort required to learn the replacement notation may be substantial enough to consider using the more difficult system from the beginning. Systems that are able to support new material without radically altering the



notation are called 'extensible'. This issue of 'extensibility' has not been addressed in this project, except in passing, but is an important factor in making representational choices for learners.

All the experiments and studies have shown how learners use diagrams to learn elementary logic. The qualitative outcomes of these studies should contribute to better logic teaching, better use of diagrams in teaching and learning and better understanding of interactive diagram design.

Diagrams are not a panacea for supporting learners with difficult subject material. However the project has increased the value of diagrams by providing some insight into their effective use.

### *8.1 Contributions.*

Educational technology is an interdisciplinary research field that draws from several other disciplines. The methods of educational technology research include controlled condition studies found in experimental psychology, focus group techniques found in marketing, as well as surveys and questionnaires found in the social sciences and other methods used in computer software design. Although the methods of study in the field may be eclectic, the criteria of success is more stable. The success of educational technology research may be best measured by the improvement of technologies for teaching and learning.

This project used several methods, including a card sorting method never used before. Results may be valuable to a readership from several communities. The development and design of interactive graphical systems will benefit from this work. The results may not generalize to every topic area. More artistic and emotive use of imagery as well as less formal graphical systems may not benefit from the results so much as formal topics. However, using graphical representations in logic teaching is similar to their use in teaching, physics, material sciences and several other disciplines.

#### *8.1.1 Empirical studies.*

The project included five studies with multiple forms of data and analysis. A small scale action research project in Chapter 5 provided qualitative results that guided later study. The bigger study ( $n = 42$ ) reported in Chapter 6, used three software applications, a pre and posttest and balanced the subject population sample for mathematics training. Three more studies with a new population sample are reported in Chapter 7. This more in-depth study with 20 subjects used a cognitive style test, a card-sorting test with a posttest interview. The final study collected video data from learners solving syllogisms with and without software



support. The complexity and depth of data increased in each study design. The research questions included in each study also increased with each new study.

The first study involved three mathematics students from a secondary school and the current author as instructor. This loosely constrained action research approach was essential in defining the parameters for the more controlled studies that followed.

The cognitive science literature showed that certain kinds of behaviours are often demonstrated by people trying to solve the syllogism. Across cultures very similar errors were presented. The literature shows no evidence that training subjects in these reasoning skills would improve performance. Most research in this area demonstrates the 'innate' tendencies of the human mind. Until the first study it was not clear that secondary school students would be able to learn this material. The value of the software tools was also unknown.

The abstract systems were all created for pedagogic purposes. However in their software implementation, very limited evaluations seemed to exist. One of the tools, the EULER system, was not created for teaching, but to clarify the algorithm in the original system.

With three learners in the room and a blackboard the dialogue was recorded to audio tape and transcribed for later analysis. Test questions were developed during dialogue and parts of the circle systems were used on the blackboard. After two hours of teaching interaction with the blackboard, learners had a short time to use the software.

The small scale study produced encouraging outcomes. Learners in the early parts of the dialogue showed signs of the predicted misconceptions from the literature and the instructor was able to identify them. At the end of the dialogue several of the misconceptions had been corrected. Learners comments on the software were positive in each case. The study showed the subjects are good candidates for improving this reasoning skill, and that the software tools would be likely to help.

The second study was more carefully controlled. Forty-two subjects were selected from two populations within a local secondary school. Half of this group studied mathematics and the other half primarily humanities subjects. The humanities subjects included, history, geography and English. The two groups were matched for abilities at their practice exam grades in the different subject areas. The two populations were split into three similar groups, making 6 conditions in a 'three by two' study set up. The three within-discipline conditions each used one of three software applications, either VENN, EULER or TARSKI'S WORLD. Additional teaching materials off-set differences in the functionality of the software.

The study led learners through instructional material and presented a pre-test. This was followed by free exploration with the software and then a post-test. Pre and posttest differences were calculated for each condition. Statistical analysis tools assessed the significance of differences in the data. Learners' posttests included worked diagram examples and final diagrams that learners created to solve problems. These diagrams proved a very rich



source of information about the errors made. Once the set of errors for a particular graphical system had been catalogued they were mapped against the disciplines.

The second study shows several new results. The Venn system turned out to be the most useful system for teaching the syllogism. Mathematics students did well with both the circle systems. The TARSKI'S WORLD system did not support significant improvement in students pre to posttest scores. The analysis of errors and misconceptions led to a catalogue of errors. Different bugs and errors were presented by the mathematics students than by the humanities students.

These results led to several questions. The two main surprises from this second study were that (a) humanities students did so badly compared to students with mathematics interests, and (b) the predictions of the information enforcement metric did not equate with the value that learners derived from the systems.

The Chapter 7 studies were designed to address these new questions. If there was another explanation for the failure of the metric to predict outcomes, then the value of the metric might be preserved, and the new variable included in a decision support system.

A cognitive style test tried to address differences in performance between the two discipline conditions. Results showed that the mathematics students were using a primarily serialist approach to solve the spy ring problem. These mathematically trained students also scored highly on the versatile style indicating that they are able to move between styles as the problem dictates. Humanities students closely correlated with a holistic approach. These results do not explain the reason for the humanities students' failure to benefit from the graphical systems used in this study, although the results are consistent with an hypothesis described in Chapter 7.

A strongly indicated reason for failure of non-mathematically trained learners, in the Chapter 6 study, is that these learners used a partly successful strategy that was dropped when they realised it could not be fully successful. Without any other active strategy, these learners scored no better than chance results in the posttest.

Results of Chapter 6 study showed that less specific systems could sometimes be better learning support tools. The Chapter 7 study addresses this issue with a new hypothesis. The Chapter 6 study data appeared to indicate that learners had more difficulty with the graphical lexicon in some systems than in others. Difficulty in perceiving the meaning of graphical elements of the lexicon is unrelated to information enforcement. The card-sorting study provides a method to assess learners' preferences for all each of the lexicon used in the graphical systems.

Results of the study showed that the more specific system also had the most preferred lexicon. This result meant that dislike of Euler's graphical conventions did not explain its lack of performance with learners in the Chapter 6 study. Several of the subjects in this card-

sorting test were later interviewed. Transcripts of these interviews illustrated learners explanations for their choices.

The last part of the Chapter 7 study collected video data from instructors training learners in syllogistic reasoning with and without software. This study revealed a new beneficial behaviour exhibited by successful learners, and resulted in a model of syllogism problem solving with diagrams.

Eight main categories of contribution illustrate the breadth of impact in the results. The main audience for the work will be those interested in using cognitive science to guide the use of technology based logic learning. However, many researchers interested in cognitive modeling, learning styles, computer based learning, logic teaching and experimental methods, may find use for the results. The methodology could be applied to a very wide range of technology based learning topics.

- A catalog of learners' bugs, misconceptions, errors, and beneficial activities using diagram systems to learn syllogistic reasoning informs a detailed learner model. The model can be used by logic teachers, logic learners, and logic learning system designers.
- A methodological approach incorporating exploratory, qualitative and quantitative data collection and analysis, provides a model for educational technology research. Innovation in the card-sorting technique provides a ranking system that will enhance qualitative studies.
- The development and application of the information enforcement principle demonstrates the possibility of applying cognitive theories of information processing to guide learning systems design. The metric was effective in predicting differences between systems with big differences in information enforcement. Researchers in cognitive disciplines may be encouraged to use implementation criteria to evaluate their theories. System designers may be encouraged to draw more from cognitive disciplines.
- Reflection on the use of intermediate representations to count system states informed the theory of graphical representation use in problem solving.
- Cognitive style analysis revealed connections between preferred disciplines of study and styles of learning. This will contribute to curriculum design and help to inform learner's study choices.
- A review of the uses of representations in learning from the learning sciences literature established a framework for empirical studies with learners.
- A detailed analysis of seven diagrammatic systems for teaching syllogistic reasoning and three software programs, reveals operational, logical and pedagogic differences in design. Logic teachers as well as designers of logic teaching software, may use these descriptions to choose software for learners. Designers and teachers may use the approach to compare software in other topics.
- The design of each study required supplementary teaching materials. These materials can be re-used by teachers of elementary logic.



### 8.1.2 Buggy and misconception model.

The study reported in Chapter 6 used learners' posttest drawings and work-scratchings to develop a model of misconceptions and errors. Analysis of the data from learners using the Venn system revealed seven major misconceptions. For five of these misconceptions, the annotations made with each rule are predictable. A set of annotations made by learners operating with the mal-rules contribute to improved diagnosis of learner errors.

Analysis of the data from learners using the Euler system also revealed seven misconceptions. These errors are more difficult to generalise for automated diagnosis. A distribution is calculated for each set of errors across the two dependent groups. Significant differences appear between the errors made by mathematics learners and those not actively studying mathematics.

Errors, misconceptions and mal-rules are useful in the development of student models, selecting instructor interventions and for providing feedback to the learner. The current model provides a detailed analysis of learners' errors with several graphical systems and the different errors shown by learners with and without mathematical backgrounds.

Informing instructors about the kinds of errors made by learners will better prepare them to decipher learner's statements more quickly in dialogue. Instructors may choose to teach the topic using learners' errors as teaching opportunities. The mal-rules and errors may need to be refined and tested before and during implementation as computer based training.

Some of the unexplained errors shown by Euler users appear to be the result of failing to properly register the middle term circles. There is little explanation of learners who fail to see all the intersections possible in the Euler circles. Pathological diagrams in all systems are unexplained by definition. As long as learners produce diagrams that defy interpretation, there will be work to do to improve the model.

The buggy model can be used for diagnosis as shown in Chapter 6. Although the results in Chapter 6 illustrate the diagrams expected from learners with different bugs, some diagrams indicate several bugs.

Sequencing the order of new assessment items can help to differentiate between competing underlying causes with similar external manifestation. Particular questions can help to differentiate competing underlying causes.

The inferences from external diagram to underlying causes are not properly tested in this project. Assessment of these inferences will reveal probabilities rather than strict implications from diagrams to errors. Technologies such as Bayesian networks and Dempster-Shafer algorithms (Petrushin & Sinita, 1993; Huang, Collins, Greer & Dobson, 1996), will be used in future work to model the diagnosis of bugs from learners' diagrams.



The sequential probability ratio test (Frick, 1992) has been applied to produce an efficient method for deriving a student model. This is described below in limitations and future work.

A discussion of the buggy and misconception model in Chapter 6 includes the criticism that this kind of approach encourages the technification of learning at the expense of helping to understand deeper problems (Laurillard, 1988). Chapter 6 argues that error analysis is necessary for the design of guided and challenging learning with interactive graphical simulations and microworlds.

### 8.1.3 A methodological approach

The methodological approach throughout the project has been analytical, empirical and exploratory. Aspects of research methods from several disciplines contribute to the processes used in this project. Phenomenographic approaches informed the interaction with learners and analysis of data in all the empirical studies. Qualitative approaches ensured relatively deep explanations of the phenomena, while comparative designs provided the environment for statistical analysis of outcomes. Qualitative and quantitative studies revealed quite unexpected outcomes in the results. These outcomes demanded new forms of explanation and new hypotheses. Hypotheses generated during the project were tested with specialised tests from the learning sciences literature, such as the 'spy-ring' test.

Analysis of the information processing principle provided a mathematically calculable approach to comparing graphical systems for teaching logic. This form of analysis involves no data. The measure of such an analysis is in how well it captures the intention of the principle.

The first unexpected result was a byproduct of trying to apply the new principle to the Venn diagram system. Counting the individual states for this diagram system proved complex but we discovered that the pattern of states had a very strong isomorphism with another, more familiar representation. The edges, vertices and faces of a wire-frame cube were mapped to areas of the Venn diagram. This discovery was a direct result of the exploratory approach. The wire-frame cube supported some of the more difficult parts of the project. The cube is an intermediate representation that simplifies calculation. This use of the wire-frame cube is an innovation that constitutes an achievement in its self and it is discussed below.

The analysis produced measures for each of the systems. The value of the principle depends on how useful each system is for learners and if differences between systems are matched by the predictions of the analysis.

Simple controlled studies evaluated relative effectiveness among the three abstract graphical systems for learning the syllogism. The studies required some adjustment since the software versions of the abstract systems included operational and pedagogic support that was different for each tool. The objective of these studies (Chapter 6) was to measure differences



between the abstract systems, not the software. The adjustments equalised some of the confounding differences between the software conditions.

A broad range of analysis and data collection techniques were used in all studies. Cognitive style tests were developed from research reported in the literature. This approach is a labor intensive test that requires substantial commitment from the learners and the researchers.

Pre and posttests were developed based on complex factors in the domain. For these tests to be genuine indicators of change, several factors were controlled such as the believability of the conclusions, and the similarity of form in pre to post test questions. Changes in pre to posttest scores inform a measure of performance differences attributable to the teaching and learning intervention.

Interactions between learners and instructors using pencil and paper, as well as software versions of the systems, were recorded on video and audio tape. This data helped to understand the dynamic activities that good learners use with interactive diagrams. Video footage of learners also contributed to the development of the buggy library and to a model of learning syllogistic reasoning that is expressed in Chapter 7 as a concept map. Learners' diagrams produced during problem solving were collected and analysed and contributed to the buggy library.

Analysis of the quantitative data consisted mostly of comparisons between controlled conditions that produce parametric test results or frequencies of error. The comparison of errors demonstrated by different discipline groups is an example of analysis to determine group allocation. This analysis showed that non-mathematicians made significantly different errors than mathematicians. The SPSS software was used to analyse this data. Transcribed audio data used a simple rubric and excerpts illustrate patterns in learners' explanations. Analysis of video sequences used a sequential technique called ESDA, together with purpose built analysis software called Timelines.

The methodological approach is appropriate for exploratory investigation in educational technology with graphical materials. The controlled quantitative data is balanced by designs that capture supportive and illustrative explanation. The method proved successful in this project and would likely serve the needs of other similar projects in educational technology research.

#### *8.1.4 The information enforcement metric.*

One of the main goals of the project was to render an abstract principle of information processing into a simple mathematical function. The specificity principle was interpreted into a ratio called the information enforcement metric. The metric was assessed by its author (Stenning, personal communication, August 29 1995), who agreed it was a reasonable



interpretation of the specificity principle. It is worth restating our definition of the metric here. The metric is an interpretation of Stenning's ideas that: (a) graphical representations are one kind of system which exhibit specificity – that is they compel specification of classes of information in contrast to systems which allow more abstraction, (b) such representations are easy to process, and (c) this specificity helps to explain why graphical techniques such as Euler's circles are so didactically effective.

The definition and constituent parts of the metric went through several iterations but in its final form is very simple. The expense of this simplicity is that several systems cannot be evaluated. Systems with graphical lexicon that can be combined and manipulated with infinitely long strings of symbols cannot be evaluated with the metric. It is only possible to assess systems with the metric, that have a limited maximum number of states.

Chapter 3 illustrated the application of the metric to a simple office indicator system. Chapter 4 described the principle more clearly and applied it to four interactive graphical systems for illustrating the syllogism.

The metric is the key part in the process of applying a theory of diagram processing to guidance for developing instructional systems. Implementing and testing a cognitive theory is one way to evaluate its plausibility. Tests of the metric appear to indicate that small differences in information enforcement are outweighed by other factors in the learning situation, for example training in other disciplines and understanding of graphical systems are factors that affect learnability as much as 'specificity' does. The results of learners using different systems with large differences in information enforcement, are consistent with the predictions from the specificity principle.

The metric represents an advance in the development of a micro-media selection system that will, with future development and improvement, be a valuable addition to instructional designers. The metric will be improved by finding and incorporating other factors involved in media selection. Further work to develop the metric is described below in the section on further work.

#### *8.1.5 Cognitive style analysis.*

Using the spy-ring cognitive style test with subjects in the two major studies in the dissertation revealed differences in learning strategies that are mirrored with learners' choice of study topics. Several recent studies, which are briefly described in Chapter 7, have indicated problems experienced by learners beginning mathematics and formal logic learning at the University level. The results show that learners who choose to study mathematics at the advanced secondary school level adopt a style of learning quite different to those learners who choose other non-mathematical study topics. Mathematics learners adopt a systematic analytical approach that is, according to Pask's interpretation of the results of their test, aided



by mechanistic memorization techniques and by encoding information without reference to the context and embeddedness of the domain. These same mathematics learners are also able to change style when appropriate, if the learning problem demands it.

Learners who have chosen not to study mathematics in the studies illustrated here, use a comprehension style and according to Pask's work tend to incorporate new materials while making semantic links with the meaning inherent in the domain. They pick up relevant cues from all available information. These learners showed no ability to shift between learning approaches and performed poorly in the learning tasks.

These results add to increased understanding of training needs for mathematics and formal reasoning learners. Previous studies, also discussed in Chapter 7, identified 'fear' as a major obstacle to developing formal reasoning skills. Learners showing problems with formal reasoning in this dissertation were from a similar population to those studied in previous studies. However the cognitive style test suggest that a more complex set of factors contribute to performance, than 'fear' alone.

#### 8.1.6 Representation use in learning science literature

The second chapter in the dissertation reviewed several descriptions of ways in which diagrams and other graphical representations are effectively used during learning and problem solving. The chapter contributes an organizing framework to study different kinds of diagrams used in several different ways. The framework is useful because it draws from many disciplines which each use terms that are difficult to differentiate. The framework may be strongest at describing kinds and uses of formal graphical systems and may therefore be most useful to designers considering development of technology based learning in formal topic areas. The framework makes one particularly prescriptive conclusion that representations should not be thought of in isolation. Learners and problem solvers are most successful, and system designs are most supportive, when translation between representations can be accommodated.

#### 8.1.7 Intermediate representations.

Applying the metric to the graphical systems turned out to be a complex and time consuming process. The limiting rules for each graphical system are open to interpretation and differences in usage. The underlying logic alters the number of system states allowable and the constraints on compatible symbol combinations are complex.

The difficulties in counting the states in the Venn system were balanced by the discovery of a useful graphical device. The structure of the Venn diagram is similar to the



structure of a wire-frame cube. The cube has a more familiar structure with names for parts such as vertices and faces which in the structure of the Venn diagram have no name.

In Chapter 4, the cube provided a support system for counting the independent states in the Venn diagram. Stages in the counting argument reflect visual patterns in the wire-frame cube.

Using another language or representation to solve a problem in a different language is not a new idea. The originality of this particular advance is the discovery of the wire-frame cube as an intermediate representation for the Venn and Carroll systems.

Several features of the cube helped make counting the numbers of states a more tractable problem. The cube has a familiar naming system. The equivalent parts of the Venn diagram have no name, making visualisation of patterns in the Venn diagram much more difficult. Both the Venn diagram and the wire-frame cube have several lines and planes of symmetry. Counting the number of states in a system is helped by these symmetries because they enable repetition and multiplication of solution steps. Combining the familiarity of parts in the cube with the enhanced symmetry meant that counting states in the Venn diagram was much easier while using the wire-frame cube as a frame of reference.

The idea and use of intermediate representations illustrate both the value of translation between multiple representations as well as the need for appropriate expressivity in the representations used. Translating Venn diagrams into wire-frame cubes had a powerful effect on understanding the mechanism and parts of the diagram.

The wire-frame cube language is isomorphic with the Venn language, but more powerful than the Venn diagram because of its familiarity, and symmetry. Chapter 7 shows the value of familiarity for learners using representations in problem solving. The card-sorting study results in Chapter 7 show that the learners' familiarity with a representation directly impact their initial understanding of the diagrams.

Cognitive science research addresses the processes involved in problem solving. The insights into to 'intermediate representations' described in this project, add to an understanding of problem solving. More study of intermediate representations in problem solving would help to categorise the value of multiple representations in human information processing.

#### *8.1.8 Support systems for syllogistic reasoning.*

Chapter 3 describes seven graphical systems for demonstrating the syllogism. The analysis revealed differences in the underlying logic and a trade-off between algorithm complexity and lexical complexity. The descriptions provide a comparison between the abstract systems and the implemented software versions.



Descriptions of the software systems illustrate an added layer of difference between each tool. The software includes more than the abstract systems described in Chapter 3. The software also includes methods for displaying the topic to the learner, for interacting with the topic and provide feedback to the user. Descriptions of the software systems are useful for logic teachers and for designers of logic teaching software. They are also valuable to designers of software for teaching other topics. The instructional strategies and feedback included in these tools may be transferable to many other areas.

There are strengths and weaknesses in the design of each system. The comparison between systems, together with indications from other parts of the thesis that multiple representations can be valuable, suggest a new design that combines aspects of some of the currently available software, into a new multiple representation software tool for teaching elementary logic.

#### *8.1.9 Teaching materials and strategies*

Each stage of the empirical studies required development of supplementary teaching materials. The initial materials provided a common level of exposure to the language of elementary logic. Working with this "Introduction to Aristotelian logic" formed the first part of the Chapter 6 study for learners. This introductory booklet has no diagrams to illustrate problems. Other materials described the function of each diagram system. One booklet was written for each system, "The Euler system and the syllogism, " describes the use of the Euler system for solving syllogism problems, "The Venn system and Aristotelian logic" describes the Venn system. Another similar booklet called, "Using Tarski's World to learn Aristotelian logic," fulfills the same function. Each booklet provides exercises and a test. Each booklet illustrates several syllogisms in increasing level of difficulty.

There are several guidelines incorporated in these instructional booklets, that could inform good software design in this area. Each system is portrayed as a process with constraints, and not a static representation of objects and relations. Learners are encouraged to create their own problems. Problems are presented in order of difficulty based on model complexity.

Using these guidelines in conjunction with the bug catalogue would support the design of improved teaching tools for this and other topics using interactive graphical representations.

#### *8.2 Limitations.*

The project work reported here was carefully planned and implemented. Well-defined goals were addressed with an appropriate methodology for the discipline. However, the mixed

methodology meant that population samples had to be small enough to make qualitative study possible and large enough to allow measurable and significant differences between conditions. Some quantitative results are not conclusive.

### *8.2.1 Limitations of the methodology*

The study methods helped to guide data collection and analysis during the investigation and application of a theory of information design. The methods carefully balance the numbers of learners and the depth of analysis, to produce results that are both statistically clear and supported by rich illustration. The exploratory design enabled new hypotheses to be created and tested on the strength of unexpected results from previous studies.

However, not all of the criteria for conclusive experimental designs are captured in the methods. The two major studies in Chapters 6 and 7 used a voluntary selection criteria for the experimental groups. This may have slightly biased the sample. The samples may not be completely representative of secondary school learners. Nevertheless, both samples were chosen in the same way so that comparisons should still be valid. The additional effort to match subjects' prior scores from relevant exams in both conditions also helped to off-set any effects of voluntary selection criteria.

The statement of exact probabilities may be exaggerated since many of the significant results are from the same experiments with the same subjects. Subjects are not strictly independent. The studies that use the same subjects in Chapter 7, are however, very different studies. Interactions between these sample should not be expected.

Another limitation in the method was discussed briefly in Chapter 6 and concerns the effort to build a catalog of errors from the error data. It is possible that modeling techniques, although empirically tested (Brown & Burton, 1978; Young & O'Shea, 1981), may not be completely defensible. Learners' behaviours are difficult to diagnose and errors turn out to be unstable. Grouping learners' work-scratchings to determine a bug, then implementing the inferred bug to show the bugs are reproducible from the model, may be methodologically problematic (Hennessy, 1994). However, the proof of any teaching system lies in its ability to support the learner. This is a more straight-forward empirical question.

The methodology used is generally appropriate for educational technology research of this kind. However, results of some studies could be reinforced by repeating them with larger more controlled groups of learners.



### *8.2.2. Limitations of the results*

Certain limitations of the results from these studies stem from the reliability of the study designs. Remarks above suggest that larger more controlled studies could reinforce the results. The reliability of implications drawn from the results also depends on the interpretation made of the theory in the initial hypothesis.

The information enforcement metric provides the beginning of a decision support system for choosing between alternative graphical representations. Many of the results in the study have illustrated other factors that affect representational choice and would need to be included in a metric. The metric relies on a simple arithmetical interpretation of the cognitive theory it is intended to characterise. It may be improved by including better understanding of complexity theory and may draw from different approaches in cognition such as the viscosity principle (Fitter & Green, 1979; Green, 1989).

The Chapter 7 study matches cognitive styles with discipline choices. This result demonstrates a correlation between a well known measure of cognitive style and mathematically trained learners. This result is consistent with the hypothesis that learners without mathematical training make use of belief cues in problem solving. However, our original hypothesis suggests that learners drop this strategy when they find it unhelpful. Tests of cognitive approaches to problem solving describe effective learners changing between strategies, but do not describe learners dropping strategies without replacing them with anything better. The results of the cognitive style test are consistent with the idea that non-mathematics trained learners behave differently than other learners, but the phenomena requires more explanation than the results are able to provide.

### *8.3 Further work.*

The work described in this thesis is preliminary investigation into the development of decision support for representation choice. The work also investigates problems in the syllogistic reasoning domain and the problems of non-mathematics trained learners. Further experiments and developments of the information enforcement metric might usefully be carried out and various extensions and related new studies are outlined below.

### 8.3.1 *An enumeration machine*

The analysis in the project that caused the most difficulty was the process of counting the states of representational systems. Counting the individual states by hand is very time consuming and prone to error. Although the discovery and introduction of the intermediate representation improved the method for at least two of the systems, the method could benefit from automation.

Representational systems have constraints such as Venn's maximum of twelve regions, eight arcs, a star and shading. One line of development would produce code to check the findings and add to the more complex versions of the metric described below.

An automated decision support system for representation selection would allow faster and more accurate calculation of system states. This process has begun with a calculation of the states in the Venn system. The code, written in the "C" programming language, appears in the appendices of the thesis.

### 8.3.2 *Carroll and other software*

Prototype software has been developed to replicate the Carroll paper-based system. The tool was built so that it mirrored the functionality of the Venn system. The specificity count for the Carroll system (Chapter 4), is greater than for Venn. Consequently Carroll should not have supported learning any better than Venn or Euler. The empirical studies generated a different result than predicted by the initial specificity principle. It would be useful to continue this implementation and to run another condition with this software and a comparable population sample. This software written in HyperCard could form the basis for implementation of a much improved system for teaching the syllogism. The system would use the general results of the thesis including bug categories, difficulty ratings, believability factors, and alternative representations.

### 8.3.3 *Towards an improved metric*

The thesis investigates criteria for selecting optimal interactive graphical representations and strategies for the learner. This project has addressed several factors and it is now possible to make a well-informed attempt at describing the relationship between them. The description involves several variables that have not been consolidated yet in any formal metric. New work



would involve integrating these factors into a better metric and testing the results in other domains. Factors that are important to diagram system choice include,

- size of the domain,
- expression of the system,
- method of achieving system expression
- complexity of algorithm
- semantics of domain misconceptions
- opportunities for cognitive conflict
- opportunities for translation
- expressive distance between representations

Studies reported in this project have shown that, the ratio of information presented in a teaching and learning system, compared to the capacity of the representation to present information, yields a factor that predicts the benefit learners derive from using the system.

The size of the domain sets a lower bound beneath which a representational system would be inadequate for the task. A representational system that could only convey a part of the domain would have to be rejected as an option. The results of evaluating the specificity principle leave the importance of the metric open to interpretation.

Algorithm complexity almost certainly affects the benefit of these systems. One clue to this factor is in the development of the Euler system (Stenning, & Inder, 1995), from the older versions (Ceraso & Provitera, 1971; *see also* Chapter 3). The reduction in the number of system states in the new version is achieved at the expense of a more complex (although shorter), algorithm for solving the problems.

Since the specificity metric only counts superficial differences in visible representations in a system, and does not evaluate the complexity of steps needed to use those diagrams to solve the problems, this factor has been ignored until now. Failure of the information enforcement metric to predict differences in outcomes between Euler and Venn may be due to neglect of algorithm complexity. The SIGNAL project (Stenning & Oberlander, 1994) emphasised that representations are 'systems' not 'static illustrations' but it is unclear if descriptions of specificity should be modified to account for algorithm complexity.

The complexity of the algorithm is at least one interacting variable that caused a deviation in the data between the two representations systems with similar specificity. It would be very difficult to prove this affect, since changes in representational systems almost necessarily mean changes in several variables at once. The study needed to find the value of this variable is analytic rather than empirical. The effort is in finding the right analytic tool. Some possible tools are described later, from complexity in computer science.

The methods used to achieve expression must also affect a system's impact on the learner. The indicator for the office door used in Chapter 4 illustrates how the same product of permutations can be achieved with dissimilar tokens and a small range of values or with a



smaller number of tokens and larger range of variable values. This difference is described in complexity by 'dimensional' and 'cardinal' accounts. Systems that encourage familiarity by reusing tokens will be more simple to use than those that use several unmemorable tokens. This phenomena is illustrated in the value of the wire-frame cube in Chapter 4. Translating the problem of permuting states in Venn to the problem of recognising patterns in the cube, allowed the familiar symmetries of the cube to emerge and simplify the problem.

Another factor involves the opportunities that each system provides for the learner to challenge typical misconceptions. The dynamic aspects of the EULER system are very compelling in this respect. Euler animates the constraining relationships between circles that are free to move in any direction without violating those constraints. This appears to provide opportunities for cognitive conflict that can lead to learning. The studies have not conclusively shown this factor, and show that steps in registering diagrams are more complex than those in the VENN system.

Providing learners with the opportunity to translate between representations is a valuable instructional strategy (Laurillard, 1988). All the systems described here use several representational languages. TARSKI'S WORLD has microworlds and predicate calculus, diagramming systems have labels, spatial locations and objects. In each case the problem begins as a sentence in natural English and must be translated at least once.

It is likely that there are optimal measures for the relationship between these mixed representations. The difficulty for the learner in moving from one representational language to another involves dis-embedding the concept from the language enough to require new constructions to be made in the new language. The dis-embedding should not be so great to make connections difficult with the new language. If the difference between the languages is optimal, the 'sense' of the concept will be retained and expanded during translation. Determining optimal difference in specificity between the languages used may provide the answer. Since we have described the value of specificity as relative to the size of the domain, it is also likely that this *difference* will be relative to the size of the domain too.

The metric created and developed in Chapters 2 and 4 is the beginning of an ideal micro-media selection model (Dijkstra, 1997). The specificity principle contributes to a functional metric for media selection but more information is necessary to make the metric reliable. Chapter 4 showed how to use an interpretation of specificity for systems of representation with a finite number of states. However most representations cannot be considered in this way because the presence of unlimited strings in a system make the number of states effectively unlimited. An improvement to the interpretation could include complex state machines augmented transition networks and other devices to analyse nonfinite systems.

Other relevant principles of information presentation should be considered such as viscosity (Fitter & Green, 1979; Green, 1989). This principle is related to specificity and is concerned with the overhead a system causes a learner. Instead of predicting the overhead



through a cardinal measure of ability to express domain information, viscosity is concerned with users' difficulties in manipulating symbols to solve a problem.

It is not immediately possible to see how this principle might be quantified or formalized. However, Green (personal communication, January 18, 1996) has indicated that the search for a metric from this principle might be of benefit both in developing the principle and in leading towards a model of modality or notation choice.

Developing and testing new metrics could constitute a long and difficult research program that would have to draw from many different research disciplines.

#### *8.3.4 A curriculum for learning from graphics.*

The thesis has shown there are general skills in using graphical systems and that learners' abilities vary widely. Learners' skills with graphical systems are related to discipline choice and to cognitive style. These factors are probably not permanent or inflexible and the skills are likely to improve with teaching.

A substantial effort in a mixed interdisciplinary project would begin to build a curriculum that could improve general skills in the use of graphical representations in learning and problem solving.

A curriculum would certainly include guidance on choosing the right level of formality in a graphical representation for the right kind of problem (Goel, 1995; Cox, 1996). Techniques for exploring interactive graphical systems could include exploration of limiting cases and studying extreme situations. Building and inventing new representations might help learners understand the requirements of graphical systems. Describing ways that graphical systems can mislead the learner would provide a basis for critical interpretation of graphical systems.

A curriculum for improving use of graphical systems could draw from similar initiatives in text literacy. Tests of "graphicy" would measure learners' level of ability at interpreting, working with and even creating graphical representations for communication.

The interdisciplinary nature of the task and the current level of understanding indicate such a project would require substantial time, effort and coordination.

#### *8.3.5 An adaptive syllogistic tutor.*

The dissertation provides several suggestions for improved systems for learning syllogistic reasoning. One approach to improvement involves the modified use of adaptive testing algorithms for sequencing problems for the learner. Implementing the design below might involve a programmer for several months.

Qualitative analysis of the teaching and learning interactions during the study reported in Chapter 5 led to implications for automating the sequencing of problems for the learner. Data from other studies in syllogistic reasoning (Johnson-Laird & Bara, 1984; Johnson-Laird & Byrne, 1991) provide values for difficulty. Methods of adaptive testing (Huang, Collins, Greer & Dobson, 1996) need data in a different format.

This project provides an innovative method for using known data and would implement a system of adaptive testing for syllogistic reasoning. An adaptive testing algorithm such as sequential probability ratio testing (Frick, 1992; Huang, Collins, Greer & Dobson, 1996), can determine the mastery and nonmastery status of a learner from limited information about that learner. The outcome of a sequential probability ratio test or SPRT, is a classification of a learner as either master or nonmaster of the content domain (Wainer, 1990). The classification is based on the numbers of questions that the learner has correctly or incorrectly answered. For each question a master is more likely to give a correct answer. Individuals who are eventually classified as masters of a topic may have incorrectly answered several questions during the adaptive interaction. The test eventually reaches a classification based on prescribed probabilities from empirical study.

The classification of the SPRT algorithm is based on an implementation of Bayesian probability. An SPRT ends when the likelihood that the learner belongs to either the master or nonmaster population is sufficiently high. The algorithm is based on four conditional probabilities, a probability ratio and three decision rules. The four conditional probabilities indicate how likely a master or nonmaster will correctly or incorrectly answer a question. They are based on historical data that has to be derived from empirical evaluation.

The data for this domain is presented as the results of cross cultural study by Johnson-Laird and Bara (1984) and by Johnson-Laird and Byrne (1991). For a single question, learners who have mastered the domain or subset, may have a mean score of say, 85%. Learners from the nonmastered group may present mean scores of 40%. The probabilities can be shown as:

*Rule 1A: Prob(Correct/Master) = .85;*  
*Rule 1B: Prob(Incorrect/Master) = .15;*  
*Rule 2A: Prob(Correct/Nonmaster) = .40, and,*  
*Rule 2B: Prob(Incorrect/Nonmaster) = .60.*

The probability ratio or PR is used to represent the likelihood that the student belongs to a class either master or nonmaster and is calculated as below.



$$PR = \left\{ \frac{P_{om}P_m^r(1-P_m)^w}{P_{on}P_n^r(1-P_n)^w} \right\}$$

Where:

- PR = probability ratio
- P<sub>om</sub> = prior probability of mastery
- P<sub>on</sub> = prior probability of nonmastery
- P<sub>m</sub> = probability of correct response for a master (determined by rule 1a)
- P<sub>n</sub> = probability of correct response for a nonmaster (determined by rule 2a)
- r = number of correct answers so far
- w = number of wrong answers so far

There are three decision rules that divide the PR space into three ranges, (a) master, (b) nonmaster, and (c) undecided. The decision rules use two parameters  $\alpha$  and  $\beta$  provided by the test designer.

- SPRT decision rule 1:* If  $PR \geq (1-\beta)/\alpha$ , classify the student as master and stop test.
- SPRT decision rule 2.* If  $PR \leq \beta/(1-\alpha)$ , classify the student as nonmaster and stop test.
- SPRT decision rule 3.* If  $\beta/(1-\alpha) < PR < (1-\beta)/\alpha$ , select another question to continue test.

The value of  $\alpha$  depends on the test designer's willingness to misclassify a nonmaster as a master which is essentially equivalent to the probability of making a false mastery decision. The value of  $\beta$  is the probability of making a false nonmastery decision.

The basic limitation of this SPRT algorithm is the requirement for the historical data for values in rules 1A through 2B:  $P(C/M)$ ,  $P(I/M)$ ,  $P(C/N)$ , and  $P(I/N)$ . We are currently carrying out analysis that will overcome this limitation and use more commonly available data in a more efficient way. Since the total number of correct answers from the student population is equal to the number of correct answers from master learners and number of correct answers from nonmaster learners we may construct a general formula:

$$M*\alpha + N*\beta = \Delta$$

Where,

- M is the fraction of learners that the author wants to consider as masters*
- N is the fraction of learners that the author want to consider as non-masters*
- $\alpha$  is the probability that a master correctly answers the question  $P(C/M)$ ,*

$\beta$  is the probability that a nonmaster correctly answers the question  $P(C/N)$   
 $\Delta$  is the probability that a student in the student population correctly answers the question.

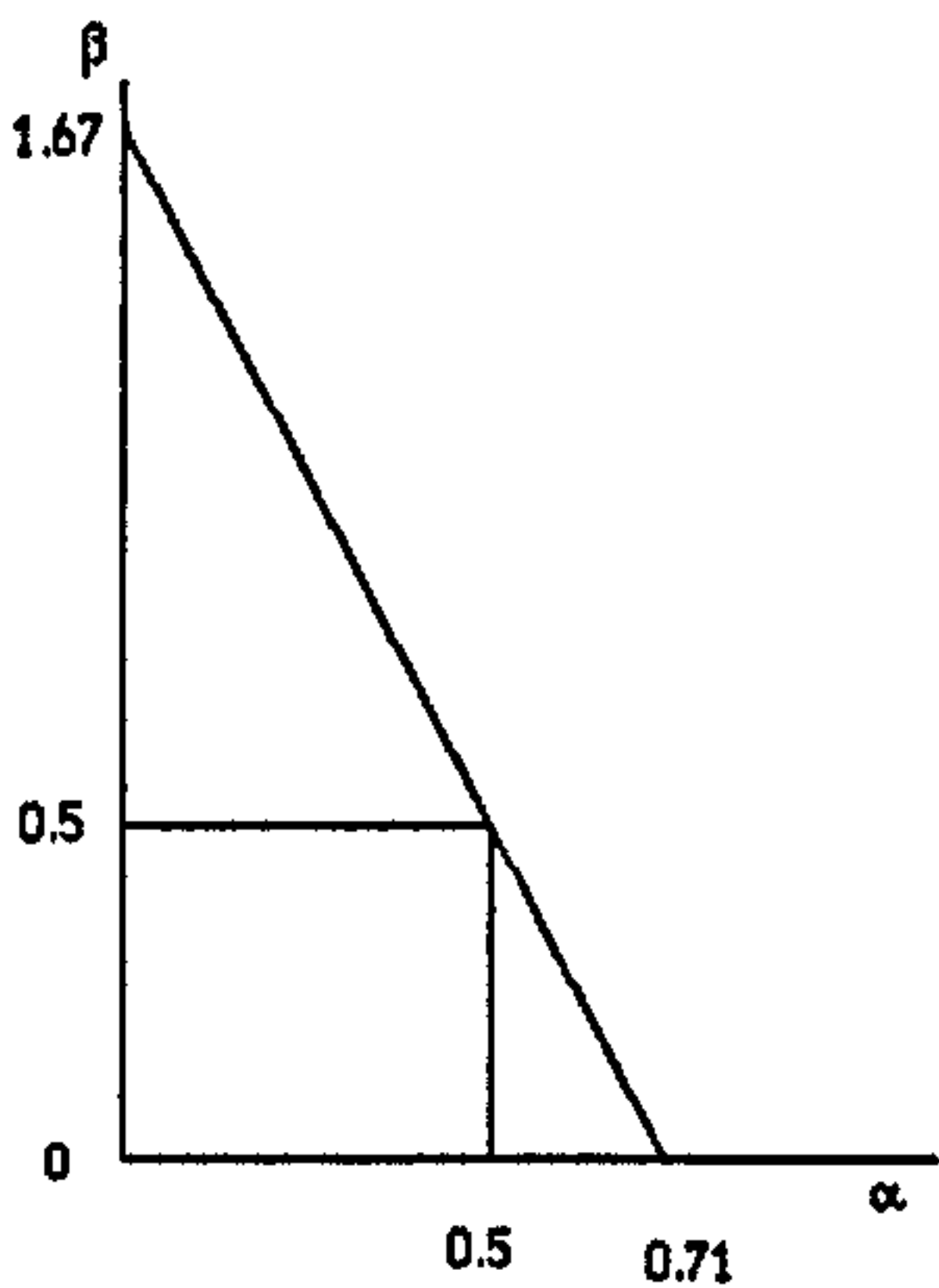
If we require the test author to specify the proportion of learners who will pass the test and be designated master we make the author specify the value of  $P(C/M)$ . Once we have  $P(C/M)$ , then,  $P(C/N)$  can be obtained from  $P(C/N) = 1 - P(C/M)$ . For example, if the course author expects that 70% of learners would pass the test then after instantiating M and N we could be left with the following:

$$0.7*\alpha + 0.3*\beta = \Delta$$

Since our goal is to obtain values of  $\alpha$  and  $\beta$  for each question so that SPRT will be usable, we have assumed that for each question we have the known percentage of learners who correctly answered the question (Johnson-Laird & Byrne, 1992; Chapter 3). In the main we can use this value as  $\Delta$ . In this way we have derived a further variable. So for the question where there is a 50% chance of a population getting it right, then we can have the formula below.

$$0.7*\alpha + 0.3*\beta = 0.5$$

This is a linear function which can be depicted as below.



Note that when  $\alpha < \Delta$  (0.5 in this example),  $\alpha < \beta$ , which means that a nonmaster would have a better chance to correctly answer the question than a master. The likely values for a and b are on the line segment between these points  $(0.5, 0.5)$  and  $(0.71, 0)$ . Without any further information it is reasonable to assume the middle point  $(0.605, 0.25)$  as values of  $\alpha$  ( $P(C/M)$ ) and  $\beta$  ( $P(C/NM)$ ). Thus we have the values for the following probabilities.

$P(C/M) = 0.605$	$P(I/M) = 1 - P(C/M) = .395$
$P(C/NM) = 0.25$	$P(I/NM) = 1 - P(C/NM) = .75$

This principle can clearly be extended to cover cases for every mean return of correct scores for a population. For example the syllogism All Bs are As, All Bs are Cs  $\Rightarrow$  Some As



are Cs has expected mean correct response rate of 19%. Our algorithm would yield the following function.

$$0.7*\alpha + 0.3*\beta = 0.19$$

Which would yield the values for each of the four parameters as follows.

$$P(C/M) = 0.23$$

$$P(I/M) = 1 - P(C/M) = 0.77$$

$$P(C/NM) = .095$$

$$P(I/NM) = 1 - P(C/NM) = 0.99$$

These probabilities are predictive enough to provide realistic feedback to a learner and to determine the end point of an adaptive test. This work will continue and will be empirically tested against a set of questions delivered in the normal way. Individuals using this test will be categorised as masters and nonmasters. The same questions and answers will be fed to the algorithm since answers by each individual are known at this point and the comparison of classification ability will be made.

### 8.3.6 Investigating direct manipulation interactive graphics

Microworlds add value by exposing and testing learners' prior conceptions and provide favourable conditions for cognitive conflict. Resolution of this conflict requires the learner to explicitly make the causal relations between the components of their own view of the domain model (Draper *et al.* 1992). Direct manipulation of simulations that are representative of the domain are meant to provide the space for conflict and resolution. Learners' errors often persist even after semistructured and guided discovery with the simulations. Learners' inability to create a runnable model is often blamed for these errors.

The investigations in this thesis indicate that there are other factors that impact on the success of direct manipulation interfaces for learning. Several studies could be done to investigate the impact of these factors on designs for simulation and other direct manipulation interfaces.

One such factor lies in the dynamic aspects of the graphics. As multimedia becomes more 'active', there is increasing pressure to add 'activity' even when it does not contribute to benefit for the learner. The dynamic nature of the representations of EULER have been underexamined in this work. Access to the constraints in a simulation by means of graphical manipulation may be more important than formal measures of expressivity.

Claire O'Malley (1990) distinguished between semantic directness and articulatory directness. An interface is semantically indirect if the user has to engage in a lot of planning and problem solving in using the system to perform some task. Articulatory directness refers to the extent to which the user's understanding of the meaning of expressions maps onto the form of expression required by the system. Her comments are directed at interface designs in general. The message applies to interactive diagrams more strongly.



Studies to investigate and test the effects of these differences could select or develop interfaces with articulatory and semantically direct manipulation and compare learning outcomes. Results would establish the benefit to the learner in creating interaction with diagrams that explicitly represent the important and difficult concepts with interactive graphics.

Such a project would require some time to clarify the exact design for the study and the methods used for data collection and analysis. Differences in interaction with semantically and articulatory direct manipulation graphical systems may be best shown from video recordings of those interactions.

### *8.3.7 Improving knowledge worker skills with formal reasoning training.*

Given the importance of the critical thinking skills taught in philosophical logic courses and the difficulty a large number of learners have mastering them, we have been investigating the use of software programs to provide practice. The project involves three strategies to provide a workable solution.

- *Create a database of graded exercises and test questions in early logic accessible through the WWW. The PAN—Tutor provides a delivery system based on adaptive testing (Huang, Collins, Greer & Dobson, 1996) and will also be used to track student performance and assign exercises to learners based on achievements and needs.*
- *Adapt existing programs intended for individual use such as TARSKI'S WORLD (Barwise & Etchemendy, 1991), Symlog, and others so that learners can practice these exercises and receive immediate feedback on their work. This may be done within a WWW text based conferencing system.*
- *Test the effectiveness of the systems with learners of different ages and backgrounds as an integrated part of the current teaching activity at The University of Calgary.*

Skills developed in studying first-order logic and argument are recognised as important tools for people who work in knowledge-based economies. Research shows that teaching reasoning skills can lead to higher performance in problem solving. Learners who acquire skills in philosophical argument become better at critical reading and writing and at assessing knowledge in a broad range of areas. They become skilled at analysing and organising complex texts and tasks (Stenning & Oberlander, 1994). Transfer of these skills has been shown for students attempting standardised tests such as the graduate record exam (Stenning, Cox & Oberlander, 1995). These abilities are central to lifelong learning and adapting to change and particularly relevant in computing technology. Several groups of students; adult learners, students with poor mathematics backgrounds, and students with English as a second language, need supportive environments and extra exercises in order to master skills of formal logic. Translating written arguments into logical notation, interpreting statements written in formal notation, recognising good and bad arguments, and constructing formal arguments, all constitute areas of difficulty. Learners need exercises that match their



individual learning levels and quick feedback on exercises so that they do not repeat their mistakes.

Assigning and grading exercises using the WWW allows learners to practice and learn whenever and wherever suits their learning style and their schedule of other activities. This aspect is especially important for adult learners and others with family and work commitments. Evaluation of computer based programs for learning logic has also shown that certain programs can help alleviate the common problem of 'fear of formal reasoning' (Fung, O'Shea, Goldson, Reeves & Bornat, 1994). The immediate outcomes of this project will be:

- *Increased achievement of mastery among learners who face barriers to learning logic skills: learners with poor mathematics skills, English as a second language learners, and learners with a 'fear of formal reasoning'.*
- *Greater flexibility which will facilitate learning among adult learners and other nontraditional groups who may have diverse responsibilities outside the university.*
- *An increase in the amount of time professors have to spend with struggling learners through the automation of mechanical grading tasks.*

The initial target audience of the project consists of the students who currently take logic courses at The University of Calgary. These students are primarily composed of philosophy majors and computer science majors. Many of the Philosophy majors especially those involved in the co-op education program are contemplating careers in the civil service and in business where critical thinking and information analysis skills taught in logic are particularly valuable. Computer science majors find logic skills an essential component of their work. These skills are valuable for all knowledge workers and are particularly important for programming and software development. Finally, business management and social science students constitute a significant minority of logic students. These students, like their philosophy counterparts are entering careers that place a high emphasis on critical thinking, problem analysis, and information assessment. There are increasing numbers of returning adult learners for whom English is a second language.

This project is novel in the juxtaposition of the graphical microworld of TARSKI'S WORLD and a conferencing system to support that collaboration. The results of this project will be to increase the performance of management and technical workers and to increase the availability of suitable employees capable of productive effort in a knowledge economy.

#### 8.3.8 Re-framing and re-answering the thesis question.

The original question of the thesis was framed, "What knowledge and expertise can the cognitive sciences reveal about the use of diagrams in teaching and learning, and how might that knowledge be used to design better computer based learning." After this long and detailed research project we are now in a position to reframe this original question as, "How



do factors known to impact usability and learnability of diagrams, together with factors associated with the mapping of domain semantics with manipulation control, affect learning with diagrams?" This re-framed question depends on a range of factors of interactive diagrams that together with *specificity* appear to affect their benefit to learning and problem solving. Such factors were found through analysis of both *surface* and *mechanical* aspects of diagram systems and by noticing interactions that address learners' prior conceptions. We suggest an empirical approach for acquiring learner conceptions and for matching these with diagram manipulations that allow the user to directly articulate the semantics of those conceptions. This is offered as a rational approach to improve the pedagogic benefit for particular diagrams and as a useful tool for designing learner-friendly computer-based diagrams.

In new work a framework will be developed from data collected in realistic teaching and learning situations that will link the semantics of interactive graphical representations with their pedagogic effectiveness for different topics and different learners. The framework will refer to empirical results in representational specificity (Dobson, 1998b; Stenning & Oberlander, 1995), algorithm complexity, illustrative affordances (Dobson, 1999; Laurillard, 1990) and appropriate task-representation matches (Green, Petre, & Ballamy, 1991). To guide selection of multiple representation strategies, the framework will refer to pedagogic elaboration using qualitative to quantitative sequencing, progressive access to the underlying model (De Jong et al., 1998), relative representational redundancy (Ainsworth, Bibby, & Wood, 1998), and to increasing the complexity of the underlying model. The empirical approach will yield a mapping between accurately described diagram constraints (Wang, 1995) and learners conceptual difficulties in a domain. Dialogue collected in realistic teaching and learning situations with a range of practitioner groups in appropriate domains will be analyzed with qualitative analysis tools such as ATLAS.ti. From this data we expect to capture a schema for coordinating multiple representations in learning that will lead to clear indications for learning science and instructional design practice.

The main purpose of the work is to develop and extend a framework around the links between the semantics of representations and their pedagogic effectiveness for different topics and different learners. The framework will draw from previous work with graphical learning systems (De Jong et al., 1998) and will be extended and continually tested through a systematic program of studies in genuine teaching and learning situations. We anticipate the benefit from the framework will be increased by the integral role of the practitioner partners who will direct the study designs to emphasize realistic learning situations. The project will draw from the best practices in the complementary use of data collection in practice oriented research with framework development and application.

The research is driven by an emerging framework that deals with representations in learning. Several complex and interacting factors are important in designing effective



interactive graphical interfaces to simulations of physical systems and mathematical domains. Recent work by a task force of the European Science Foundation on learning with multiple representations produced a preliminary framework that identifies many of these factors (Van Someren, Boshuizen, Reimann, & de Jong, 1998). The framework consists of some clear empirically tested indicators developed in small laboratory studies and some conjecture based on analysis and projection from study results. Effective media designs appear to depend on the breadth of the domain of instruction and the capacity of any representation to communicate (Dobson, 1998b; Stenning & Oberlander, 1995). However, similarly expressive representations may have varying learnability because of the method used to achieve their expression. Transformations within a representation system may require processes, or algorithms, that affect the difficulty of using the representation (Dobson, 1999). The representations may yield different affordances for illustrating and creating cognitive conflict around the prevalent learner misconceptions in the domain (Dobson, 1999; Laurillard, 1990). Different representations for a single topic may afford better support for different tasks, and mismatching the task for the representation can impede user performance (Green, Petre, & Ballamy, 1991). Representational choice in media design is further complicated because learners rarely rely on single representations, and often must translate among many different surface languages. The idea of 'specificity' for single systems (Stenning & Oberlander, 1995) describes how some languages, particularly graphical languages, may be pedagogically useful because they are limited in expressive capacity. This idea is partly scaled up to learning environments with multiple representations by the idea of 'relative representational redundancy' (Ainsworth, Bibby, & Wood, 1998). Translation among representational languages may not be automatic and often requires instructional support (Conlon, 1998). For some very simple concepts the translation strategy may not yield added insight for the learner. However, competent translation is often seen as an educational goal that can illustrate learners' understanding of the key concepts in a field (White & Fredericksen, 1990), and may lead to more flexible application of complex skills and knowledge (Spiro & Jehng, 1990). Furthermore, the process of translation effectively reduces the expressive capacity of a system of languages. The intersection of communication with two languages will be less than the union of the two languages, making the system more specific and potentially more effective (Dobson, 1998b).

Designers of learning media have also to consider the ordering of representations based on difficulty and pedagogic elaboration. Transitions from one representation to another may be guided by sequencing rules; qualitative to quantitative, access to the underlying model, and increased complexity of the underlying model. For each sequencing guideline, the links among representations must be explicitly shown (Van Someren et al., 1998) (p119), and a detailed analysis of the knowledge involved may also be needed. Designing interactive media for learning therefore requires coordination of media representations so that useful sequences



of opportunity for interaction and cognitive conflict are provided to the learner. However, technology based learning media also provides possibilities that were not previously available to the designer. In these cases the research provides little guidance and progress does not keep up with media development. Interactive graphical representations are a relatively new family of instructional tools (Dobson, 1998a). Much of the work in multiple representations applies directly to these new media but their novelty suggests a need for new guidelines. A few recent studies using dynamic displays, animated graphics (Rieber, 1990) and law encoding diagrams (Cheng, 1998) begin to address these affordances but differences even between these media have not been factored into the studies.

Analysis of the semantics of diagrams illustrates the possibility of accurately describing the constraining factors of manipulable graphics (Wang, 1995). A framework which effectively demonstrates the mapping between diagram constraints and the important semantics of a topic would help designers build interactive graphical systems that address learners' difficulties with a topic. A goal of this modeling work involves producing guidelines for designing interactive learning environments that support cognitive conflict with the known conceptual difficulties in the field. In logic learning with diagrams we identified a complex catalogue of errors and misconceptions from learners developing understanding of syllogistic reasoning (Dobson, 1999). The phenomenographic approach to needs-assessment has produced similar outcome spaces in many topic areas. We also identified specific problems of using diagrams that may be similar across all domains. Together these learner problems build a model of errors and misconceptions in using diagrams to learn reasoning skills. Taken separately, a catalogue of errors and misconceptions of learners using diagrams to learn about any discipline may have wide applicability.

To understand how learners use these instructional tools the research methods will build on the controlled condition studies reported in the field. These have often required learners to assimilate quite simple concepts during short periods of study. The teaching and learning research program occurs in a practice-oriented field. The tools under consideration and the resulting improvements in learning outcomes may benefit from more practical feedback. Improvements in learning outcomes will be the ultimate judge of success, but while the tools and techniques are developed, reliable data can be best provided by exploratory mixed data collection and analysis techniques. Just as learners rarely use single representations for learning, so they also rarely use representations without a complex dialogue surrounding the representation. Learning conversations are quite different to social conversations (Lee, Dineen, & McKendree, in press), and some work has specifically considered the tutorial dialogue around static graphical representations and the sketching process (Goldschmidt, 1991). This project will examine real dialogue between learners and between teachers and learners demonstrating interactive semantically constrained diagrams to each other. Participants in the dialogue will include instructors with topic expertise and an understanding



of teaching strategy. We expect to infer useful representation coordination strategies from this dialogue as well as new insights into the processes of learning and reasoning with interactive diagram systems. Learners' manipulation of the diagrams will be captured for analysis in synchrony with this dialogue. Several data capture and analysis tools such as ATLAS.ti and MACSHAPA will be useful for coding the dialogue and for representing the data in time related schema. The ATLAS.ti tool allows considerable support for qualitative analysis and has recently been enhanced with function for reviewing video.

At least three practitioner groups will participate throughout the study. Each group includes developers and users of software tools for learning. Each of the tools emphasize multiple interactive graphical representations. The three groups provide diversity in the developer and practitioner group; the domains involved, the level of learners' education for which the systems are being developed, and the techniques of interaction built into the learning environments. Tools that will be studied are built to support learning in: early and intermediate logic, system modeling, geology, meteorology and veterinary science. The educational levels span the primary, secondary and post-secondary system. The instructional environments and methods of interaction include; simulated microworlds built from abstract objects such as blocks and tetrahedra; immersive environments designed to simulate real worlds, and abstract graphical systems built from notations that are not analogous with the objects they represent.

The project will combine expertise in the learning processes involved in multiple representation use with instructor and developer practitioners. Narrative analysis of tutorial dialogue will be vital in guiding analysis of learning interactions with the graphical systems. From this data we expect to begin to capture a schema for coordinating (planning, manipulating, experimenting) multiple representations in learning. We expect to produce a clear model of learning with multiple interactive graphical representations that will lead to indications for instructional media design. By working with practitioners we expect to understand more about integrating these technology tools into the elementary, secondary and post-secondary system. We also expect practitioners will develop a researcher-practitioner approach to their continuing work and will continue to benefit from the tools and methods used during the study.

#### *8.4 Implications for educational technology.*

This dissertation has demonstrated the value of interdisciplinary research in supporting successful educational technology development. An integrated and mixed methodology was used. This was derived partly from study practices in human computer interaction, partly from quantitative methods of experimental design and partly from phenomenographic approaches. Implications for the design of computer based learning systems, include the implementation of models of learner's misconceptions, and the beginnings of a decision support system related to the information enforcement principle. The work has highlighted

several unanswered questions about the best use of dynamic and static graphic representations. The application of cognitive theory to educational technology has been a challenge. Certain stages of the research can be considered as successes in themselves, including; capturing the phenomena described by the theory, interpreting the theory with some accuracy and focusing on a tractable and yet convincing problem. The research encountered more problems than answers. This may not be unusual, however, the research methods meant that the outcomes from all the studies were valuable, if not always as expected. The long-term goals of educational technology as a research discipline remain tied to the development of our understanding of cognition and to the adoption of eclectic research methods. Future work will make opportunistic selection among alternative approaches. Exploratory and analytical methods have the best chance to contribute to better teaching and learning software and better experiences for the learners. If teaching and learning systems are to achieve a better match with learners' competencies, our continued investigation of cognition will continue to have profound impact.





## References

- Ainsworth, S. E., Bibby, P. A., & Wood, D. J. (1998). Costs and Benefits of Multi-Representational Learning Environments. In M. Van Someren, H. P. A. Boshuizen, P. Reimann, & T. de Jong (Eds.), *Learning with Multiple Representations* (pp. 120-137). Amsterdam: Pergamon.
- Allwein, G., & Barwise, J. (Eds.). (1996). *Logical reasoning with diagrams*. Oxford: Oxford University Press.
- Anderson, M., & McCartney, R. (1996). *Diagrammatic reasoning and cases*: Initial report. Hartford, Connecticut: University of Hartford.
- Anderson, J. (1978). Arguments concerning representations for mental imagery. *Psychological Review*, 85 (4), 249-277.
- Anderson, J. R., Conrad, F. G., & Corbett, A. T. (1989). Skill acquisition and the LISP Tutor. *Cognitive Science*, 13, 467-506.
- Anderson, J. R., Corbett, A. T., Koedinger, K., & Pelletier, R. (1995). Cognitive tutors: Lessons learned. *The Journal of Learning Sciences*, 4, 167-207.
- Arcand, J. F. (1994). HIT : Un système pour l'évaluation et la création d'une interface humain-ordinateur intelligente [HIT: A system for evaluation and design of human computer interfaces]. *Proceedings of 12th International Ergonomics Association.*, Vol. IV, (pp. 380-385).
- Aristotle. (1983). *Prior analytics* (H. Tredennik. Trans.). Cambridge, MA: Harvard University Press. (Original work published 336 BC)
- Arshad, F.N., & Kelleher, G. (1993). SOLA\*: Students on-line advisor. *International Journal of Man-Machine Studies*, 38, 281-312.
- Ausubel, D.P. (1963). *The psychology of meaningful verbal learning*. New York: Grune & Stratton.
- Baeker, R. M., Grudin, J., Buxton, W.A.S., & Greenberg, S. (Eds.) (1995). *Readings in Human Computer Interaction: Toward the year 2000 (2nd Edition)*. New York: Morgan Kaufman.
- Baggett, P. (1987). Learning a procedure from multimedia instructions: The effects of film and practice. *Applied Cognitive Psychology*, 1, 183-195.
- Barwise, J., & Etchemendy, J. (1990). Visual information and valid reasoning. In W. Zimmerman & S. Cunningham (Eds.), *Visualisation in Teaching and Learning Mathematics* (pp. 9-24). Washington DC: Mathematical Association of America.
- Barwise, J., & Etchemendy, J. (1991). *The language of first order logic: Including the program Tarski's World 3.0 (second edition)* CSLI Lecture notes, no. 23. Stanford: CSLI Press.



- Barwise, J., & Etchemendy, J. (1995). *Heterogeneous logic*. In J. Glasgow, N. H. Narayanan, & B. Chandrasekaran. (Eds.). *Diagrammatic reasoning: Cognitive and computational perspectives* (pp. 211-234). Cambridge, Mass: AAAI and MIT Press.
- Barwise, J. & Perry, J. (1983). *Situations and attitudes*. Cambridge, MA: The MIT Press.
- Bates, A. W. (1995). *Technology, open learning and distance education*. London: Routledge.
- Behr, M.J., Lesh, R., Post, T.R. & Silver, E.A. (1983). Rational number concepts. In Lesh, R. & Landau, M. (Eds.), *The acquisition of mathematics concepts and processes* (pp. 7-44). Orlando, FL: Academic Press.
- Begg, I., & Denny, J. (1969). Empirical reconciliation of atmosphere and conversion interpretations of syllogistic reasoning. *Journal of Experimental Psychology*, 81, 351-354.
- Bernsen, N. O. (1993). *Foundations of multimodal representations: A taxonomy of pure modalities* (Working Papers in Cognitive Science & HCI). Roskilde University, Denmark, Centre for Cognitive Informatics.
- Bernsen, N.O., & Bertels A. (1994). *A methodology for mapping information from task domains to interactive modalities* (Working Papers in Cognitive Science and HCI). Roskilde University, Denmark, Centre for Cognitive Informatics.
- Belnap D. N. Jr., & Steel, T. B. (1976). *The logic of questions and answers*. New Haven and London: Yale University Press.
- Bertin, J. (1983). *Semiology of graphics: Diagrams, networks and maps*. Wisconsin: Univ. of Wisconsin Press.
- Boden, M. (1988). *Computational models of mind*. Brighton, UK: Harvester.
- Bovet, D. P., & Crescenzi, P. (1994). *Introduction to the theory of complexity*. Prentice International Series in Computer Science. Oxford: Prentice Hall.
- Bricken, M., & Byrne, C. (1993). Students in Virtual Reality: A Pilot Study. In Alen Wexelblat (Ed.), *Virtual reality: Applications and explorations* (pp. 246-273). San Diego, CA: Academic Press.
- Brna, P. (1996). Can't see the words for the tree: Interpretation and graphical representations. In T. R. Addis (Ed.), *IEE Colloquium Digest, 51. Colloquium on Thinking with Diagrams*. Savoy Place, London UK.
- Brown, J.S., & Burton, R.R. (1978). Diagnostic models for procedural bugs in basic mathematical skills. *Cognitive Science*, 2, 155-92.
- Brown, J. S., Collins, A., & Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher*, 17, 32-41.
- Bruner, J. S. (1960). *The process of education*. New York: Random House.
- Bruner, J. S. (1966). *Towards a theory of instruction*. Cambridge, MA: Belknap Press of Harvard University Press.
- Bultena, B. and Ruskey, F. 1998 Venn Diagrams with Few Vertices *The Electronic Journal of Combinatorics*. Volume 5(1), 1998

- Carr, B., & Goldstein, I. (1977). *Overlay: A theory of modeling for computer-aided instruction* (Technical Report AI Memo 406). Cambridge, MA, USA.: Massachusetts Institute of Technology.
- Carroll, L. (1958). *Symbolic logic and game of logic*. New York: Dover Publications. [Originally published in single volume as Symbolic Logic in 1896].
- Ceraso, J., & Provitera, A. (1971). Sources of error in syllogistic reasoning. *Cognitive Psychology*, 2, 400-410.
- Chapman, L.J., & Chapman, J.P. (1959). Atmosphere effect re-examined. *Journal of Experimental Psychology*, 58, 220-226.
- Charniak, E., & McDermott, D. (1984). *Introduction to artificial intelligence*. Reading, Mass: Addison Wesley.
- Chandrasekaran, B., Glasgow, J., & Narayanan, N. H. (1995) (Eds.). *Diagrammatic reasoning: Cognitive and computational perspectives*. Cambridge, Mass: AAAI and MIT Press.
- Chandrasekaran, B., & Narayanan, H. N. (1993). Perceptual representation and reasoning. In L. S. Sterling (Eds.), *Intelligent Systems: Concepts and Applications* (pp. 67-76). New York and London: Plenum Press.
- Chandrasekeran, B., Narayanan, H. N., & Iwasaki, Y. (1993). Reasoning with diagrammatic representations: A report on the spring symposium. *AI Magazine*, 14(2), 49-56.
- Cheng, P. C. H. (1994). An empirical investigation of law encoding diagrams for instruction. *Proceedings of the 16th Annual Conference of the Cognitive Science Society* (pp. 171-176). Hillsdale, NJ: Lawrence Erlbaum.
- Cheng, P. C. H. (1998). Interactive law encoding diagrams for learning and instruction. In M. W. Dobson (Ed.), *Learning and Instruction (Special edition)*. Amsterdam: Elsevier.
- Clark, J. M., & Paivio, A. (1991). Dual coding theory and education. *Educational Psychology Review*, 3(3), 149-170.
- Collins, A., Brown, J. S., & Newman, S. E. (1989). Cognitive apprenticeship: Teaching the crafts of reading, writing and mathematics. In L. B. Resnick (Ed.), *Knowing, learning and instruction: Essays in honor of Robert Glaser* (pp. 453-494). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Conlon, T. (1998). Alternatives to rules for Knowledge-Based Modelling. *Instructional Science, Kluwer Academic*.
- Cox, R. (1996). *Analytical reasoning with multiple external representations*. Unpublished doctoral dissertation, Department of Artificial Intelligence, University of Edinburgh, Scotland.
- Cox, R., & Brna, P. (1995). Supporting the use of external representations in problem solving: The need for flexible learning environments. *Journal of Artificial Intelligence in Education*, 6, 2-3.
- Dalton, O.M. (1825). A portable dial in the form of a book with figures derived from Raymond Lull. *Archeologia*, 74 (second series), 89-102.
- Daniel, J.S. (1996). Mega-Universities and knowledge media: Technology strategies for higher education. London: Kogan Page



- Davidson, P. I. (1994). The systems dynamics approach to computer-based management learning environments: Implications and their implementations in Powersim. In J.D.W. Morecroft & J.D. Sterman (Eds.), *Modeling for learning organizations*, (pp. 301-316). Portland: Productivity Press.
- De Jong, T., Ainsworth, S., Dobson, M. W., Van der Hulst, A., Levonen, J., Reimann, P., Sime, J. A., Someren, M., Spada, H., & Swaak, J. (1998). Acquiring knowledge in science and math: The use of multiple representations in technology based learning environments. In M. Van Someren, P. Reimann, P. A. Boshuizen, & T. de Jong (Eds.), *Learning with Multiple Representations*. Amsterdam: Pergamon.
- Dijkstra, S. (1997). Educational technology and media. In S. Dijkstra, N. Seel, F. Schott, and R. D. Tennyson (Eds.), *Instructional design: International perspectives*, Vol. II. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Dijkstra, S., Seel, N., Schott, F., & Tennyson, R. D. (Eds.) (1997). *Instructional design: International perspectives*, Vol. II. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Dobson, M. W. (1989). *Ready about: A tutoring system for teaching RYRU rules for avoidance of collision at sea*. Unpublished MSc. dissertation, University of Sussex, UK.
- Dobson, M. W. (1993a). Towards a model for a collaborative courseware authoring system. *Journal of Computer Assisted Learning*, 9, 34-50.
- Dobson, M.W. (1993b), Knowledge mediation, graphical representation and ITS. In T. Conlon (Ed.), *AI tools and the classroom: Theory into practice. In the proceedings of seventh international PEG conference (Vol. 2)*, 613-620.
- Dobson, M.W. (1994a). Learning through and by visualisation: A case of intermedia translation. In S.D. Franklin, A.R. Stubberud., & L.P Wiedeman (Eds.), *University education uses of visualization in scientific computing* (pp. 77-94). North-Holland, Amsterdam: Elsevier Science. ISBN: 0-444-81543-0
- Dobson, M. W. (1994b). Towards an analysis of visual media in learning: A study in improving syllogistic reasoning. In T. Ottmann & I. Tomek (Eds.), *Educational Multimedia and Hypermedia*, (pp. 153-159). AACE ISBN: 1-880094-10-X
- Dobson, M.W. (1996). Learning with specific and non-specific graphic systems. In T. R. Addis (Ed.), *IEE Colloquium Digest, 51. Colloquium on thinking with diagrams*. Savoy Place, London UK.
- Dobson, M.W., & McCracken, J. (1996). Resolution of conversational breakdowns with the distance learning toolkit. In P. Brna, A. Paiva & J. Self (Eds.), *Proceedings of the European conference on artificial intelligence in education*. ISBN: 972-8288-37-9.
- Dobson, M. W., & McCracken, J. (1997). Problem-based learning for evaluating multi-media in science technology and society. In T. Muldner & T. C. Reeves (Eds.), *Proceedings of Educational Multimedia and Hypermedia, 1997*. Calgary, Canada.
- Dobson, M. W. (Ed.). (1998a). *Learning and Instruction (special edition)*. Amsterdam: Pergamon.
- Dobson, M. W. (1998b). Toward decision support for multiple representations in teaching early logic. In M. Van Someren, H. P. A. Boshuizen, P. Reimann, & T. de Jong (Eds.), *Learning with Multiple Representations* (pp. 87-102). Amsterdam: Pergamon.
- Draper, S., Mohammed, R., Byard, M., Driver, R., Hartley, R., Mallen, C., Twigger, D., O'Malley, C., Hennessey, S., O'Shea, T. M. M., Scanlon, E., & Spensley, F. (1992).



- Conceptual change in science* (CALRG Technical report 123). Milton Keynes, UK: Open University, Institute of Educational Technology.
- Du Boulay, B., O'Shea, T., & Monk, J. (1981). The blackbox inside the glass box: Presenting computing concepts to novices. *International Journal of Man - Machine Studies*, 14, 237 - 249.
- Duffy, T. M., Lowyck, J. & Jonnasen, H. H. (1993). Designing environments for constructivist learning. Berlin: Springer Verlag.
- Edholm, P. (1981). Boundaries in the radiographic image: General principles for perception of boundaries and their application to the image. *Acta Radiol.* 22, 457.
- Edholm, P. (1993). How the visual system interprets images. Panel presentation. In N.D. Gershon (Chair), *How to lie and confuse with visualization*. In electronic proceedings of SIGGRAPH-93, 1-6 August, Anaheim, CA.
- Elsom-Cook. M. T. (1990). *Guided discovery tutoring*. London, UK: Paul Chapman Publishers.
- Entwistle, N. J., & Ramsden, P. (1983). *Understanding student learning*. London: Croom Helm.
- Euler, L. (1772). *Lettres a une princesse d'Allemagne, Vol. 2. Sur divers sujets de physique et de philosophie*, Letters 102-108. [Letters to a princess of all the Russias, Vol. 2. On subjects of science and philosophy]. Basel: Birkhauser.
- Fitter, M., & Green, T.R.G. (1979). When do diagrams make good computer languages? *International Journal of Man-Machine Studies*, 11, 235-261.
- Forbus, K. (1983) Qualitative reasoning about space and motion. In D. Gentner & A. Stevens (Eds.), *Mental Models*. Hillsdale, MJ: Erlbaum.
- Frick, T. (1992). Computerized adaptive mastery tests as expert systems. *Journal of Educational Computing Research*, 8(2), 187-213.
- Fung, P., O'Shea, T., Goldson, D., Reeves, S., & Bornat, R. (1994). Why computer science students find formal reasoning frightening. *Journal of Computer Assisted Learning*, 10, 240-250.
- Funt, B. V. (1995). Problem solving with diagrammatic representations. In B. Chandrasekaran, J. Glasgow & N. H. Narayanan (Eds.), *Diagrammatic Reasoning: Cognitive and computational perspectives* (pp. 33-69). Cambridge Mass: AAAI and MIT Press.
- Furnas, G. W. (1990). Formal models of imaginal deduction. *Proceedings of the twelfth annual conference of the cognitives science society* (pp. 662-669). Hillsdale, NJ: Lawrence Erlbaum.
- Furnas, G. W. (1991). New graphical reasoning models for understanding graphical interfaces. In *Human Factors in Computing Systems, CHI'91. Conference proceedings (ACM)*, 71-78.
- Gardner, M. (1958). *Logic machines and diagrams* (2nd ed.). New York: McGraw-Hill Book Company.
- Gagne, R. M., & Briggs, L. J. (1979). *Principles of instructional design* (2nd ed.). New York: Holt, Rinehart and Winston.



- Gentner, D., & Stevens, A.L. (Eds.) (1983). *Mental Models*. Hillsdale, New Jersey: Lawrence Erlbaum Associates.
- Gilbert, J. K., Watts, M. D., & Osborne, R. J. (1985). Eliciting student views using an interview about instances technique. In L. H. T. West & A. L. Pines (Eds.), *Cognitive structure and conceptual change* (pp. 11-27). Orlando, FL: Academic Press.
- Glasgow, J. (1993). The imagery debate revisited: A computational perspective. *Computational Intelligence*, 9 (4), 309-333.
- Glasgow, J., & Papadias, D. (1995). Computational imagery. In B. Chandrasekaran, J. Glasgow & N. Hari Narayanan (Eds.), *Diagrammatic Reasoning: Cognitive and computational perspectives* (pp. 435-480). AAAI and MIT Press Cambridge Mass. Also in *Cognitive Science*, 16 (3), 355-94
- Goel, V. (1992). Ill-structured diagrams for ill-structured problems. In *Proceedings of the fourteenth annual conference of the cognitive science society*. Bloomington, Indiana: Lawrence Erlbaum.
- Goel, V. (1995). *Sketches of thought*. Cambridge, Mass: MIT Press.
- Goldin, G. A. (1987). Cognitive representational systems for mathematical problem solving. In C. Janvier. (Ed.), *Problems of representation in the teaching and learning of mathematics* (pp. 45-89). Hillsdale, New Jersey: Lawrence Erlbaum Associates.
- Goldstone, D., & Reeves, S. (1992). *Using programs to teach logic to computer scientists*. London, UK: University of London, Department of Computer Science, Queen Mary and Westfield College.
- Green, T. R. G. (1989). Cognitive dimensions of notations. In A. Sutcliffe & L. Macaulay (Eds.), *People and Computers, 5, HCI 1989. Proceedings of the fifth conference of the British computer society* (pp. 443-459). Cambridge: CUP.
- Green, T. R. G. (1996). Response to question at conference. *IEE Colloquium Digest, 51. Colloquium on thinking with diagrams*. Savoy Place, London UK.
- Gruber, T. R. (1993). A translation approach to portable ontology specifications. *Knowledge Acquisition* 5 (2) 199-220.
- Gutwin, C., Jones, M., Brackett, P., & Adolphe, K. (1995). Bringing ITS to the marketplace: A successful experiment in minimalist design. In H. Maurer (Ed.), *Educational Multimedia and Hypermedia, 1995: Proceedings of Ed-Media 1995*. Charlottesville, VA: Association for the Advancement of Computing in Education.
- Hammer, E. (1992). Reasoning with sentences and diagrams. *The Notre Dame Journal of Formal Logic*, 35, 73 - 87.
- Hammer, E. (1995). *Logic and visual information*. Stanford: CSLI Publications (also appears in, *Studies in Logic, Language and Information*. The European Association for Logic, Language and Information. ISBN 1-881526-87-9).
- Harel, D. (1992). On visual formalisms. *Communications of the ACM*, 13 (5), 514-530.
- Harrison, B. L. (1995). Designing multimedia tools for social and interactional data collection and analysis. In P. Thomas (Ed.), *The social and interactional dimensions of Human-Computer Interfaces* (pp. 204-239). Cambridge: Cambridge University Press. ISBN 0-521-45302-X.



- Hartson, H. R., & Grey, P. D. (1992). Temporal aspects of tasks in the user action notation. *Human Computer Interaction*, 7, 1-45.
- Hauptman, A.G, & McAvinny, P. (1993). Gestures with speech for graphics manipulation. *Intl. J. Man-Machine Studies*, 38, 231-249.
- Huang, S., Collins, J., Greer, J., & Dobson, M. (1996, April). *An algorithm for content-balanced adaptive testing*. (Internal report available from THE LEARNING COMMONS, 2500 University Drive NW, Calgary, Canada).
- Janvier, C. (Ed.). (1987). *Problems of representation in the teaching and learning of mathematics*. Hillsdale, New Jersey: Lawrence Erlbaum Associates.
- Johnson-Laird, P. N. (1983). *Mental models: Towards a cognitive science of language, inference and consciousness*. Cambridge: Cambridge University Press.
- Johnson-Laird, P. N. (1988). A taxonomy of thinking. In R. Sternberg & E. Smith (Eds.), *The Psychology of human thought*. Cambridge: CUP.
- Johnson-Laird, P.N., & Bara, B. (1984). Syllogistic inference. *Cognition*, 16, 1-61.
- Johnson-Laird, P.N., & Byrne, R. M. J. (1991). *Deduction*. Hillsdale, NJ. & Hove, Sussex, UK: Lawrence Erlbaum.
- Jonassen, D. H. (1991). Objectivism vs. constructivism: Do we need a new philosophical paradigm? *Educational Technology: Research and Development*, 39(3), 5-14.
- Jonassen, D.H. (1992). Evaluating constructivistic learning. In T.M. Duffy & D.H. Jonassen (Eds.), *Constructivism and the technology of instruction: A conversation*. (pp. 137-149). Hillsdale NJ: Lawrence Erlbaum Associates.
- Khuwaja, R., Evens, M., Rovick, A., Michael, J., & Patel, V. (1996). A model of tutoring based on the behaviour of effective human tutors. In C. Frasson & G. Gauthier (Eds.), *Intelligent Tutoring Systems. Proceedings of third international conference, ITS 1996* (pp. 130-138). Berlin, Germany: Springer Verlag.
- Koedinger, K.R., & Anderson, J.R. (1994). Reifying implicit planning in Geometry: Guidelines for model-based intelligent tutoring system design. In S.P. Lajoie & S.J. Derry (Eds.), *Computers as Cognitive Tools* (pp. 15-46). Hillsdale, NJ: Erlbaum.
- Kolb, D. A. (1984). *Experiential learning: Experience as the source of learning and development*. New Jersey: Prentice Hall.
- Kulpa, Z. (1995). Diagrammatic representation and reasoning. *Machine Graphics and Vision*, 3 (1-2), 77-103.
- Larkin, J. H. (1983). The role of problem representations in physics. In D. Gentner & A.L. Stevens (Eds.), *Mental Models* (pp. 75-98). Hillsdale, New Jersey: Lawrence Erlbaum Associates.
- Larkin, J., & Simon, H.A. (1987). Why a Diagram is (sometimes) worth ten thousand words. *Cognitive Science*, 11, 65-99.
- Laurillard, D. (Ed.). (1987). *Interactive media: working methods and practical applications*. New York: John Wiley and Sons.
- Laurillard, D. (1988). The pedagogical limitations of generative student models. *Instructional Science*, 17, 235-250. Dordrecht, NL: Kluwer Academic Publishers.



- Laurillard, D. (1991). Mediating the message: Television programme design and students' understanding. *Instructional Science*, 20, 3-23. Dordrecht: Kluwer Academic Publishers.
- Laurillard, D. (1993). *Rethinking university teaching: A framework for the effective use of educational technology*. London: Routledge.
- Laurillard, D. (1990). *Phenomenographic research and the design of diagnostic strategies for adaptive tutoring systems* (CITE Report No. 124). Milton Keynes: The Open University
- Lee, J., Dineen, F., & McKendree, J. (in press). Supporting student discussion: It isn't just talk. *Education and Information Technology, IFIP TC3 Official Journal*, to appear.
- Levonen, J. J. (1996). The skills of graphing: A replication with cognitive and cross-national perspectives. (Internal report available from LRDC, University of Pittsburgh, 3939 O'Hara Street, Pittsburgh).
- Levonen, J. J., & Lesgold, A.M. (1993). Visual reasoning in a coached practice environment: Visual meaning making in Sherlock. In R. Cox, M. Petre, P. Brna & J. Lee (Eds.), *Proceedings of the workshop on Graphical Representations, Reasoning and Communication, World Conference on Artificial Intelligence in Education* (pp. 23 -27). Edinburgh: University of Edinburgh.
- Lewis, M.C. (1991). Visualisation and situations. In J. Barwise, J. M. Gawron, G. Plotkin, & S. Tutiya (Eds.), *Situation theory and its applications*, Vol. 2 (pp. 553-580). Stanford: Stanford University Press.
- Lindström, B. (1980). *Forms of representation, content and learning. Göteborg Studies in Educational Sciences*, 34 (Available from ACTA UNIVERSITATIA GOTHOBURGESIS, Box 5096, Göteborg Sweden). S-402 22 ISBN 91-7346-081-8
- Loftin, R.B., Engelberg, M., & Benedetti. (1993). Applying virtual reality in education: A prototypical virtual physics laboratory. In *Proceedings of the IEEE Symposium on Research Frontiers in Virtual Reality* (pp. 67-74). Los Alamitos, CA: IEEE Society Press.
- Lybeck, L., Marton, F., Strömdahl, H., & Tullberg, A. The phenomenography of the 'mole concept' in Chemistry. In P. Ramsden (Ed.), *Improving learning: New perspectives* (pp. 81-108). London: Kogan Page.
- Maybury, M. (in press). Research in multimedia parsing and generation. *Journal of Artificial Intelligence Review: Special Issue on the Integration of Natural Language and Vision Processing Vision Processing*.
- Mackay, W. E. (1989). EVA: An experimental video annotator for symbolic analysis of video data. *SIGCHI Bulletin*, 21, 68-71.
- Marton, F., & Saljö, R. (1984). Approaches to learning. In F. Marton, D. J. Hounsell & N.J. Entwistle (Eds.), *The Experience of Learning* (pp. 36-55). Edinburgh: Scottish Academic Press.
- Marr, D., & Hildreth, E. (1980). Theory of edge detection. *Proceedings of the Royal Society of London, Series B*, 207, 187-217.
- Mason, J. (1987). What do symbols represent? In C. Janvier (Ed.). (1987). *Problems of representation in the teaching and learning of mathematics* (pp. 73-81). Hillsdale, New Jersey: Lawrence Erlbaum Associates.

- Mizoguchi, R. & Bourdeau, J (1999) Ontology aware systems in AI-Ed Research. AI Technical report 99-04 ISIR Osaka University. March 1999. Available from [www.ei.sanken.osaka-u.ac.jp/pub/all-publications.html](http://www.ei.sanken.osaka-u.ac.jp/pub/all-publications.html)
- McCarthy, J., & Hayes, P. J. (1969). Some philosophical problems from the standpoint of artificial intelligence. *Machine Intelligence*, 4, 463-504. Edinburgh: Edinburgh University Press.
- Narayanan, H. N. (1993). Imagery: Computational and cognitive perspectives. *Computational Intelligence*, 9, (4), 303-307. London, UK: Blackwell.
- Newell, A. (1990). Unified theories of cognition. Cambridge, MA: Harvard University Press.
- Newstead, S. E. (1990). Conversion in syllogistic reasoning. In K. Gilhooly, M.T.G. Keane, R. Logie & G.Erdos (Eds.), *Lines of thought: Reflections on the psychology of thinking Vol. One* (pp. 73-84). London, UK: Wiley.
- Norman, D. A. (1983). Some observations on mental models. In D. Gentner & A.L. Stevens (Eds.), *Mental models* (pp.7-14). Hillsdale, New Jersey: Lawrence Erlbaum Associates.
- Norman, D. A. (1988). *The Design of Everyday Things*. New York, USA: Basic Books.
- Norman, D. A. (1993). *Things that make us smart*. Mass, USA: Addison Wesley.
- Oakhill, J., Garnham, G., & Johnson-Laird, P.N. (1990). Belief effects in syllogistic reasoning In K. Gilhooly, M.T.G Keane, R. Logie & G.Erdos (Eds.), *Lines of thought: Reflections on the Psychology of thinking, Vol. One* (pp. 125-138). London, UK: Wiley.
- O'Malley, C. (1990). Interface issues for guided discovery learning environments. In M. T. Elsom-Cook (Ed.), *Guided discovery tutoring* (pp. 24-42). London, UK: Paul Chapman Publishers.
- Papert, S. (1993). *The children's machine: Rethinking school in the age of the computer*. NY, USA: Harper Collins.
- Parkes, A. P. (1994). A study of problem solving activities in a hypermedia representation. *Journal of Educational Multimedia and Hypermedia*, 3(2), 197-223.
- Pask, G. (1975). *Conversation, cognition and learning: A cybernetic theory and methodology*. Amsterdam, Oxford and New York: Elsevier.
- Pask, G., & Scott, B. C. E. (1972). Learning strategies and individual competence. *International Journal of Man-Machine Studies*, 4, 217-253.
- Pask, G., & Scott, B.C.E. (1975). CASTE: A system for exhibiting learning strategies and regulating uncertainties. *International Journal of Man-Machine Studies*, 5, 17-52.
- Paivio, A. (1971). *Imagery and verbal processes*. New York: Holt, Rinehart and Winston.
- Paivio, A. (1986). *Mental representation: A dual coding approach*. New York: Oxford University Press.
- Petre, M., & Green, T. R. G. (1992). Requirements of graphical notations for professional users: Electronics CAD systems as a case study. *Le Travail humain*, 55 (1), 47-70.
- Petrushin, V. A., & Sinitsa, K. M. (1993). Using probabilistic reasoning techniques for learner modeling. In P. Brna, S. Ohlsson & H. Pain (Eds.), *Artificial Intelligence in Education*, 93 (pp. 23-27). Edinburgh, Scotland: AACE.



- Pierce, C. S. (1933). The collected papers of Charles S. Peirce, Volume 4. C. Hartshorne & P. Weiss (Eds.). Cambridge: Harvard University Press.
- Pineda, L. A. (1990). *GRAPHLOG: A theory of semantics for graphics with applications to Human-Computer Interaction and CAD systems*. Unpublished doctoral thesis, University of Edinburgh.
- Presmeg, N.C. (1985). *The role of visually mediated processes in high school mathematics: A classroom investigation*. Unpublished doctoral thesis, University of Cambridge, UK.
- Pylyshyn, Z. (1981). The imagery debate: Analog media versus tacit knowledge. *Psychological Review*, 88, 16-45.
- Ramsden, P. (1988). *Improving learning: New perspectives*. London: Kogan Page.
- Regian, J., & Shebilske, W. (1992). Virtual reality: An instructional medium for visual-spatial tasks. *Journal of Communication*, 42, (4) Autumn, 136-149.
- Reigeluth, C. (1983). The elaboration theory of instruction. In C. M. Reigeluth (Ed.), *Instructional-design theories and models: An overview of their current status* (pp. 335-379). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Reiser, B. J., Kimberg, D. Y., Lovett, M. C., & Ranney, M. (1992). Knowledge representation and explanation in GIL: An intelligent tutor for programming. In J. Larkin & R. W. Chabay (Eds.), *Computer assisted instruction and intelligent tutoring systems: Shared issues and complementary approaches* (pp. 111-149). Hillsdale, NJ: Erlbaum.
- Resnick, L. B. (1982). Syntax and semantics in learning to subtract. In T. P. Carpenter, J.M Moser and T. Romberg (Eds.), *Addition and subtraction: A cognitive perspective*. New Jersey: Lawrence Erlbaum Associates.
- Revlin, R., & Leirer, V. (1978). The effect of personal biases on syllogistic reasoning: Rational decisions from personalised representations. In R. Revlin & R. E. Mayer (Eds.), *Human reasoning* (pp. 51-80). Washington, DC: Winston Wiley.
- Rich, E. (1988). *Artificial intelligence*. London: McGraw-Hill
- Rieber, L. P. (1990). Using computer animated graphics in science instruction with children. *Journal of Educational Psychology*, 82, 135-140
- Rips, L. J. (1988). Deduction. In R. J. Sternberg & E. E. Smith (Eds.), *The Psychology of human thought*. (pp. 116-152).
- Ritter, F. E., & Larkin, J. H. (1994). Developing process models as summaries of HCI action sequences. *Human-Computer Interaction*, 9, 345-383.
- Romiszowski, A. J. (1988). *The selection and use of instructional media for improved classroom teaching and for interactive, individualized instruction (2nd Ed.)*. London: Kogan Page.
- Roscelle, J., & Goldman, S. (1991). VideoNoter: A productivity tool for video analysis, *Behaviour Research Methods, Instruments and Computers*, 23, 219-224.
- Ruskey, F (1997) A Survey of Venn Diagrams. The Electronic Journal of Combinatorics. Volume 4. Available at <http://www.combinatorics.org/Surveys/ds5/VennEJC.html>
- Sanderson, P. M., James, J. M., & Seidler, K. S. (1989). SHAPA: An interactive software environment for protocol analysis. *Ergonomics*, 32, 1271-1302.

- Schmeck, R.R. (Ed.). (1988). *Styles and strategies of learning*. New York: Plenum.
- Scardamalia, M., & Bereiter, C. (1994). Computer support for knowledge-building communities. *The Journal of Learning Sciences*, 3(3), 265-283.
- Schnotz, W., Picard, E., & Hron, A. (1993). How do successful and unsuccessful learners use text and graphics? *Learning and Instruction*, 3(3), 181-199.
- Sharples, M. (1991). Computer-based tutoring of visual concepts, from novice to expert. *Journal of Computer Assisted Learning*, 7 (2), 123-132.
- Shin, S-J. (1991). A situation-theoretic analysis of valid reasoning with Venn diagrams. In J. Barwise, J. M. Gawron, G. Plotkin, & S. Tutiya (Eds.), *Situation Theory and its Applications, Part 2, CSLI Lecture Notes* (pp. 581). Chicago: Chicago University Press.
- Shimojima, A. (1996). Operational constraints in diagrammatic reasoning. Chapter 2, In, Allwein, G., & Barwise, J. (Eds.). *Logical reasoning with diagrams*, 27-48. Oxford: Oxford University Press.
- Sloman, A. (1971). Interactions between Philosophy and Artificial Intelligence: The role of intuition and non-logical reasoning in intelligence. *Artificial Intelligence*, 2, 209-225. Amsterdam: Elsevier.
- Sloman, A. (1993). Varieties of formalisms for knowledge representation. *Computational intelligence*, 9 (4), 413-423. UK: Blackwell.
- Spiro, R. J., & Jehng, J-C. (1990) Cognitive flexibility and hypertext: Theory and technology for the nonlinear and multidimensional traversals of complex subject matter. In D. Nix & R.J. Spiro (Eds.), *Cognition, education and multimedia: Exploring ideas in high technology*. Hillsdale, NJ: LEA
- Stenning, K. (1993). Depictive versus descriptive representations: A distinction, but what is the difference? *Computational Intelligence*, 9 (4), 353-355. Blackwell, UK.
- Stenning, K. (1993, December). *The modality assignment problem*. Presentation made at, *The Open University Seminar Series in Computing Cognition and Learning*, Walton Hall, Milton Keynes, UK.
- Stenning, K., Cox, R., & Oberlander, J. (1995). Contrasting the cognitive effects of graphical sentential logic teaching: Reasoning, representation and individual differences. *Language and Cognitive Processes*, 10 (3/4), 333-354.
- Stenning, K., & Inder, R. (1995). Applying semantic concepts to the media assignment problem in multi-media communication. In B. Chandrasekaran & J. Glasgow (Eds.), *Diagrammatic reasoning: Computational and cognitive perspectives on problem solving with diagrams* (pp. 303-338). Cambridge, Mass: MIT Press.
- Stenning, K., & Oberlander, J. (1994). Spatial inclusion and set membership: A case study of analogy at work. In K. J. Holyoak & J. A. Barden (Eds.), *Advances in connectionist and neural computational theory, Volume 2: Analogical connections* (pp. 447-487). Norwood New Jersey: Ablex.
- Stenning, K., & Oberlander, J. (1995). A cognitive theory of graphical and linguistic reasoning: Logic and implementation. *Cognitive Science*, 19, 97-140.
- Stenning, K. & Tobin, R. (1993). Assigning information to modalities: Comparing graphical treatments of the syllogism. In Ejerhed, E. & Lindström, S. (Eds.) *Action, Language*



- and Cognition. Proceedings of International Conference on Dynamic Semantics.* Umea, Sweden: Kluwer.
- Tabachnek, H. J. M., & Simon, H. A. (in press). Alternative representations of instructional material. In D. Peterson (Ed.), *Alternative representations: An interdisciplinary theme in cognitive science*.
- Tennant, N. (1986). The withering away of formal semantics. *Mind and Language*, 1, 302-18.
- Tennyson, R. D., & Christensen, D.L. (1988). MAIS: An intelligent learning system. In D. Jonassen (Ed.), *Instructional designs for microcomputer courseware* (pp. 247-274). New Jersey: Lawrence Erlbaum.
- Tessler, S., Iwasaki, Y., & Law, K.H. (1995). Qualitative structural analysis through mixed diagrammatic and symbolic reasoning. In B. Chandrasekaran, J. Glasgow & N. H. Narayanan (Eds.), *Diagrammatic reasoning: Cognitive and computational perspectives* (pp. 711-729). Cambridge, Mass: AAAI and MIT Press .
- Van der Pal, J. (1995). *The balance of situated action and formal instruction for learning conditional reasoning*. Doctoral dissertation. University of Twente, Netherlands. ISBN 90 90 08710 9
- Van Someren, M., Boshuizen, H. P. A., Reimann, P., & de Jong, T. (Eds.). (1998). *Learning with Multiple Representations*. Amsterdam: Pergamon.
- Vergnaud, G. (1987). Conclusion. In Janvier, C. (Ed.), *Problems of representation in the teaching and learning of mathematics* (pp. 227-232). Hillsdale, New Jersey: Lawrence Erlbaum Associates.
- Venn, J. (1894). *Symbolic Logic*. London: Macmillan.
- Venn: *A Syllogistic Logic Tutor 3.0* [computer software] (1991). Santa Barbara: Intellimation, Library for the Macintosh.
- Vygotsky, L.S. (1978). *Mind in society; The development of higher psychological processes*. Cambridge, Mass: Harvard University Press.
- Wainer, H. (1990). *Computerized adaptive testing: A primer*. Hillsdale, New Jersey: Lawrence Erlbaum Associates.
- Wang, D. (1995). *Studies on the formal semantics of pictures*. Doctoral dissertation, ILLC. Amsterdam: Institute for Logic, Language and Computation, University of Amsterdam.
- Wason, P.C., & Johnson-Laird, P.N. (1972). *Psychology of reasoning: Structure and content*. London: Batsford.
- Wenger, E. (1986). *Artificial intelligence and tutoring systems: Computational and cognitive approaches to the communication of knowledge*. Los Altos, California: Morgan and Kaufman.
- Wetherik, N., & Gilhooly, K. (1990). The figural effect in syllogistic reasoning. In K. Gilhooly, M.T.G. Keane, R. Logie, & G. Erdos (Eds.), *Lines of thought: Reflections on the psychology of thinking Vol. One* (pp. 99-108). London. Wiley.
- White, B.Y., & Fredericksen, J.R. (1990). Causal model progressions as a foundation for intelligent learning environments. *Artificial Intelligence*, 42, 99- 157.

- White, B.Y. (1993). Intermediate abstractions and causal models: A microworld-based approach to science education. In P. Brna, S. Ohlsson, & H. Pain (Eds.), *Proceedings of AI-ED 1993. World Conference on Artificial Intelligence in Education* (pp. 23-27). Edinburgh, Scotland: AACE.
- Woodworth, R.S., & Sells, S. B. (1935). An atmosphere effect in formal syllogistic reasoning. *Journal of Experimental Psychology*, 18, 451-460.
- Young, R.M., & O'Shea, T. (1981). Errors in children's subtraction. *Cognitive Science*. 5. (153-77).





## Appendices



*Data set for spy ring test.*

Learner Style Index				
Mathematics	Neutral	Operational	Comprehen	Versatile
1	1.20	4.80	2.00	0.00
2	0.90	6.20	5.20	1.70
3	0.80	3.60	2.30	2.40
4	1.60	4.90	1.20	1.90
5	1.10	5.10	4.10	2.10
6	1.30	4.10	3.00	2.20
7	2.00	5.70	3.60	1.30
8	1.50	7.10	4.90	0.00
9	0.90	3.30	7.20	2.10
10	0.70	4.80	3.90	2.50
Mean Score	1.20	4.96	3.74	1.62
Mean Score %	0.60	0.71	0.27	0.32
Humanities				
11	0.60	1.20	4.00	0.00
12	0.20	2.30	8.20	0.20
13	1.00	2.50	7.10	1.10
14	0.50	1.30	3.50	0.70
15	0.30	2.20	6.70	1.30
16	0.20	2.90	6.10	0.00
17	0.00	3.60	7.20	0.80
18	0.00	1.00	5.10	0.90
19	0.10	0.00	6.50	0.30
20	0.30	2.40	7.20	0.60
Mean Score	0.32	2.77	6.16	0.59
Mean Score %	0.16	0.28	0.44	0.12

*Note:* Data here is summarised in chapter 7, Table 25.

*Introduction to Aristotelian Logic: The syllogism.*

This brief introduction to Aristotelian logic introduces some key ideas in the area and you should make sure you understand the material here before going on the experiment. You will find out about, the syllogism, premise, conclusion, terms, quantifiers and figures. You will find a short test at the end which will help you to know how far you have learned about the subject. Syllogistic reasoning has been the study of cognitive scientists and philosophers for almost two thousand years since Aristotle and turns out to pose some quite difficult problems for human problem solvers to perform. After reading through the following few pages you will be equipped with all the definitions of concepts used in syllogistic reasoning that you will need. You should read through the materials, at your own speed and work with the learning activities at the bottom of the pages. Here is an example of a syllogism, the following pages will refer back to this example throughout. You might like to examine it now and see if you agree with the conclusion.

<p>Some philosophers are polymaths  All polymaths are intelligent  Therefore, Some philosophers are intelligent</p>
---

The next section looks at the terminology used to talk about this kind of formal reasoning. To understand syllogistic reasoning you need to have an understanding of the language used to describe the constituent parts and allowable processes in performing the reasoning skill. In this section we outline all the definitions you need to be aware of before going on to the next phase.

**THE SYLLOGISM**

A syllogism is an arrangement of two special kinds of sentences which together can generate a further sentence which tells us new information about things mentioned in the first two sentences. All three sentences are very similar in form but certain rules govern the individual parts which can exist in those sentences.

The special kinds of sentences which exist in syllogisms are called *premises*. Premises can be very short, sometimes as few as four words long , but equally, a premise can be much longer. In the box below the premises are underlined.

<p><u>Some philosophers are polymaths</u>  <u>All polymaths are intelligent</u>  Therefore, Some philosophers are intelligent</p>
---

In a syllogism the conclusion relates the terms in the premises which appear only once in each premise, the *end-terms* (see below). In the box above, the only text which is not underlined is the conclusion.

The first thing to note here is that there can only be two *terms* in any premise. Terms are the elements of the sentence which the sentence refers to. We sometimes represent the terms in a premise with the letters A, B and C. There are three terms in both of the premises added together, and so there is one term which is mentioned in both of the premises, this term is called the *middle term* and the other two terms are called the *end terms*. It is always the end terms which are the two included in the conclusion. In the box below, the terms are underlined. Make sure that you can see how the terms are arranged, how many times each term is mentioned and what kind of term each instance belongs to.

<p><u>Some philosophers</u> are <u>polymaths</u>  All <u>polymaths</u> are <u>intelligent</u>  Therefore, Some <u>philosophers</u> are <u>intelligent</u></p>
---

There are four ways to add statements of quantity about the terms used in any premise or conclusion. The terms are replaced by the letters A and B in these examples: Remember that from the example, A would stand for Philosophers, B would stand for Polymaths and C would stand for Intelligent.



All A are B  
Some A are B  
No A are B  
Some A are not B

So in our standard example, the quantifiers are underlined.

Some philosophers are polymaths  
All polymaths are intelligent  
Therefore, some philosophers are intelligent

There are specific definitions associated with each of the quantifiers.

1. A premise which uses the quantifier *All* , for example, *All As are Bs* or All Academics are Beekeepers, is true if there are *no* Academics which are *not* Beekeepers.
2. For the premise, *Some Academics are Beekeepers*, to be true, the minimum number of academics who practice beekeeping must be just one. Also When All Academics are Beekeepers, then *Some Academics are Beekeepers* is true also.
- 3 . No As are Bs is true only when there are No Academics who keep bees, and of course by implication No Bee-keepers who are academics.
4. Some As are not Bs , is true if there are some Academics who are not Beekeepers (remember that for some to be true, there need only be one which fits the category). There may of course be some Academics who are Bee-keepers, or No Academics which are Beekeepers, there may not be All Academics are Beekeepers.

There are four possible figures, which the premises of a syllogism can be written in. In each case, A and C are the end terms - occurring in conclusions - and B is the middle term, occurring in both the premises. These figures for the arrangement of the terms in the premises are as follows;

A - B	B - A	A - B	B - A
B - C	C - B	C - B	B - C

Each premise can be in one of the four moods, therefore there are 64 distinct forms in the premises. Less than half of these combinations of the premises yield valid conclusions but there will always be some kind of solution to the syllogism, even if it is simply that there is no valid conclusion.

TEST

Read the following syllogism and answer the following questions;

No fish are typists  
Some men are typists  
Therefore, some men are not fish

- Q1. What are the terms involved in the syllogism above ? How many terms are there?, and which kind of term is each one?
- Q2. Using As, Bs and Cs to replace the terms - write out the form of the premises and the conclusion in the syllogism.

*Using TARSKI'S WORLD™ to learn Aristotelian logic*

**INTRODUCTION**

The use of *TARSKI'S WORLD* for learning syllogistic reasoning skills requires the translation of syllogisms written or spoken in ordinary language to the languages used in the software. In this booklet you can find out how to do just this. There are two ways of representing syllogisms with *TARSKI'S WORLD*, the language of first order predicate calculus, of which only a small part is needed, and a simple blocks world. Both must be used at the same time for the system to be useful. Read through the following and then complete the activity in the last box which is a *self-test* with *TARSKI'S WORLD*. You will be able to refer to these materials whilst you use the software. The main things to remember when reading these pages are the following, each of which will be explained in much more detail in subsequent sections.

- (1) *Terms must be translated into independent predicates.*
- (2) *Premises must be written in the predicate calculus.*
- (3) *A blocks world is created where the premises are true, and the conclusion cannot be made false, this conclusion is then valid.*
- (4) *Convert the terms back from the predicates to the premise terms. This is now the valid conclusion.*

**PREDICATE CALCULUS**

Predicate calculus is a large language which has been developed to formalise some of the ways in which people think. It is much larger than is needed to be known for the purposes of learning syllogistic reasoning and so only a small sub-set is shown here. You should remember from the tutorial booklet that there are only four possible forms of premises, they are listed again below. This makes the job of learning the predicate calculus equivalents much easier. You should realise that you are only learning a very small part of the calculus, and all you have to do to solve syllogisms with *TARSKI'S WORLD* is to substitute the As and the Bs for the predicates you decide to use in the world you create.

Replacing terms with A & B	Predicate calculus equivalent	From sentence window
All A are B	$\forall x (A(x) \Rightarrow B(x))$	$\forall x (\text{Backof}(x) \Rightarrow \text{Cube}(x))$
Some A are B	$\exists x (A(x) \wedge B(x))$	$\exists x (\text{Backof}(x) \wedge \text{Tet}(x))$
No A are B	$\forall x (A(x) \Rightarrow \neg B(x))$	$\forall x (\text{Tet}(x) \Rightarrow \neg \text{Frontof}(x))$
Some A are not B	$\exists x (A(x) \wedge \neg B(x))$	$\exists x (\text{Cube}(x) \wedge \neg \text{Backof}(x))$

**TARSKI'S WORLD AND THE SYLLOGISM**

Here is shown the stages of translation for these premises, into premises which may be easily used in *TARSKI'S WORLD*.

*Some composers play violin*  
*No violinists play drums*

1. Convert the terms into geometric relations which can be shown in the *TARSKI'S WORLD* program. In this case we will use the relations as shown, *i.e.* Composers can be translated to Large, Violinists can be translated to Tetrahedron and Drummers can be translated to Frontof (x,a), where 'a' is an arbitrary object, and 'x' is the tetrahedra which replaced violinists. The only restriction on what terms can be used in your representation, is that they should be *independent* of each other. If something is a cube, it cannot be a Tetrahedron whether the deduction allows it to be or not. It is therefore



essential not to use shapes as different predicates, or  $\text{Backof}(a,b)$  as one predicate and  $\text{Frontof}(a,b)$  as another within the same premise.

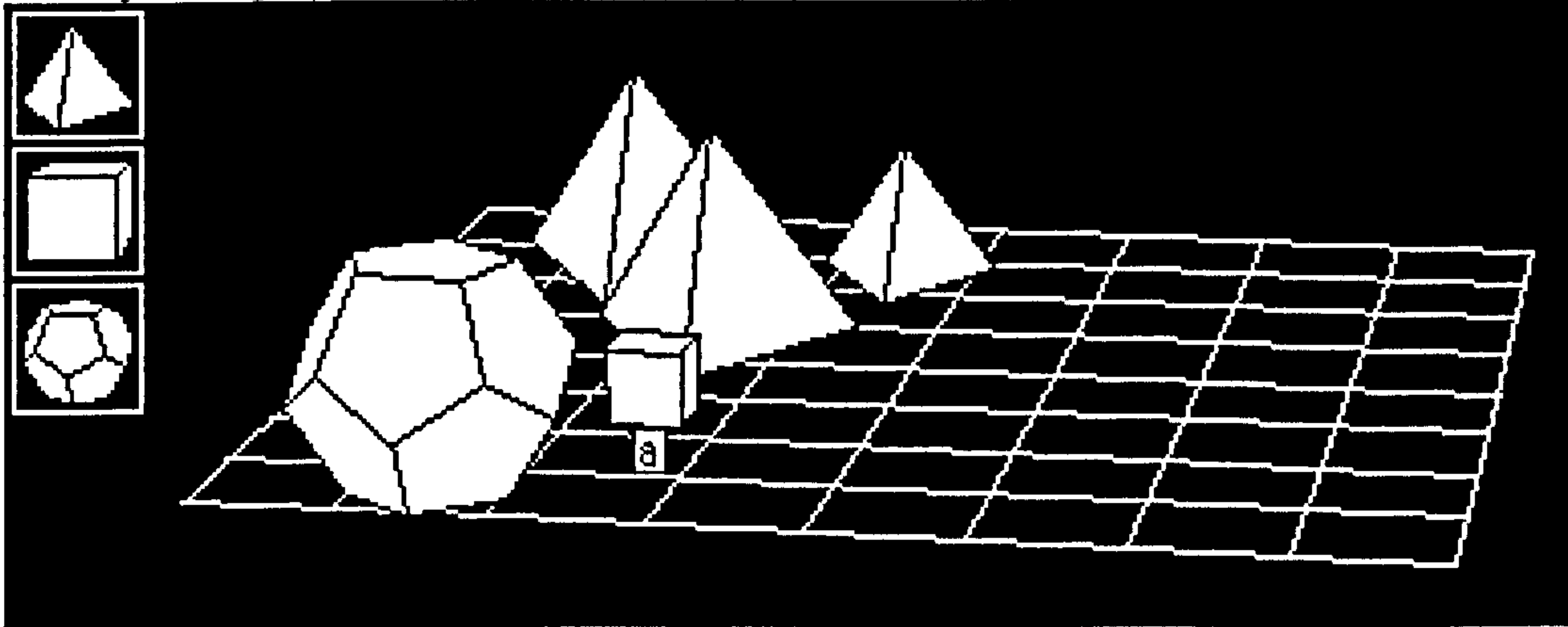
This will yield the following;

$\exists x (\text{Large}(x) \wedge \text{Tet}(x))$	Some large things are tetrahedrons
$\forall x (\text{Tet}(x) \Rightarrow \neg \text{Frontof}(x,a))$	No Tetrahedrons are in front of 'a'

2. Construct a diagram using *TARSKI'S WORLD* world module which makes the premises true, the following section of a world created in *TARSKI'S WORLD* world module is an example of this. Check this out against the sentences to see if you think both premises are true.
3. Now choose the conclusion which you think is valid. You can do this from the original premises and then convert the solution to the geometrical equivalent, or, you can go straight into the geometrical sentences. Remember, the conclusion must be one of the forms listed in the table earlier in this booklet, and relate the A and C terms. To help in this example we have listed the potential conclusions as lines in the demonstration that follows.
4. Now try to make the conclusion false by altering the diagram, but whilst keeping the premises true. If you fail to make the conclusion false, then it must be a valid conclusion.
5. In this case; Some Large things are not in front of 'a' which translates to;  $\exists x (\text{Large}(x) \wedge \neg \text{Frontof}(x,a))$  is the valid conclusion, and translates back to the original syllogism to; Some Composers don't play drums.

### EXAMPLE ONE

Some composers play violin, No violinists play drums



$\exists x (\text{Large}(x) \wedge \text{Tet}(x))$

*True: This Statement is true of the world built in TARSKI'S WORLD. It can be checked in the verification window.*

*(Some Large things are Tetrahedra)*

*[Replacing Composers with large objects and Violinists with Tetrahedra and maintaining the form Some As are Bs (I), yields the first premise]*

$\forall x (\text{Tet}(x) \Rightarrow \neg \text{Frontof}(x,a))$

*True: This Statement is true of the world built in TARSKI'S WORLD. It can be checked in the verification window.*

*(No Tetrahedrons are in front of 'a')*

*[Again replacing violinists with Tetrahedra and Drummers with objects in front of 'a', maintaining the form of the original second premise, No Bs are Cs (O), yields this premise in TARSKI'S WORLD]*

$\forall x (\text{Large}(x) \Rightarrow \neg \text{Frontof}(x,a))$

*False: This Statement is FALSE of the world built in TARSKI'S WORLD. It can be checked in the verification window.*

*(No large things are in front of 'a')*

*This statement cannot therefore be a valid conclusion from the first and second premises.*

$\forall x (\text{large}(x) \Rightarrow \text{Frontof}(x,a))$

*False: This Statement is FALSE of the world built in TARSKI'S WORLD. It can be checked in the verification window.*

*(All large things are in front of 'a')*

*This statement cannot therefore be a valid conclusion from the first and second premises.*



$\exists x (\text{Large}(x) \wedge \neg \text{Frontof}(x,a))$

*True: This Statement is TRUE of the world built in TARSKI'S WORLD. It can be checked in the verification window.*

*(Some large things are not in front of 'a')  
As much as you try to change the world represented , as long as the premises remain true - this statement will also be true . It is therefore a valid conclusion.*

$\exists x (\text{Large}(x) \Rightarrow \text{Frontof}(x,a))$

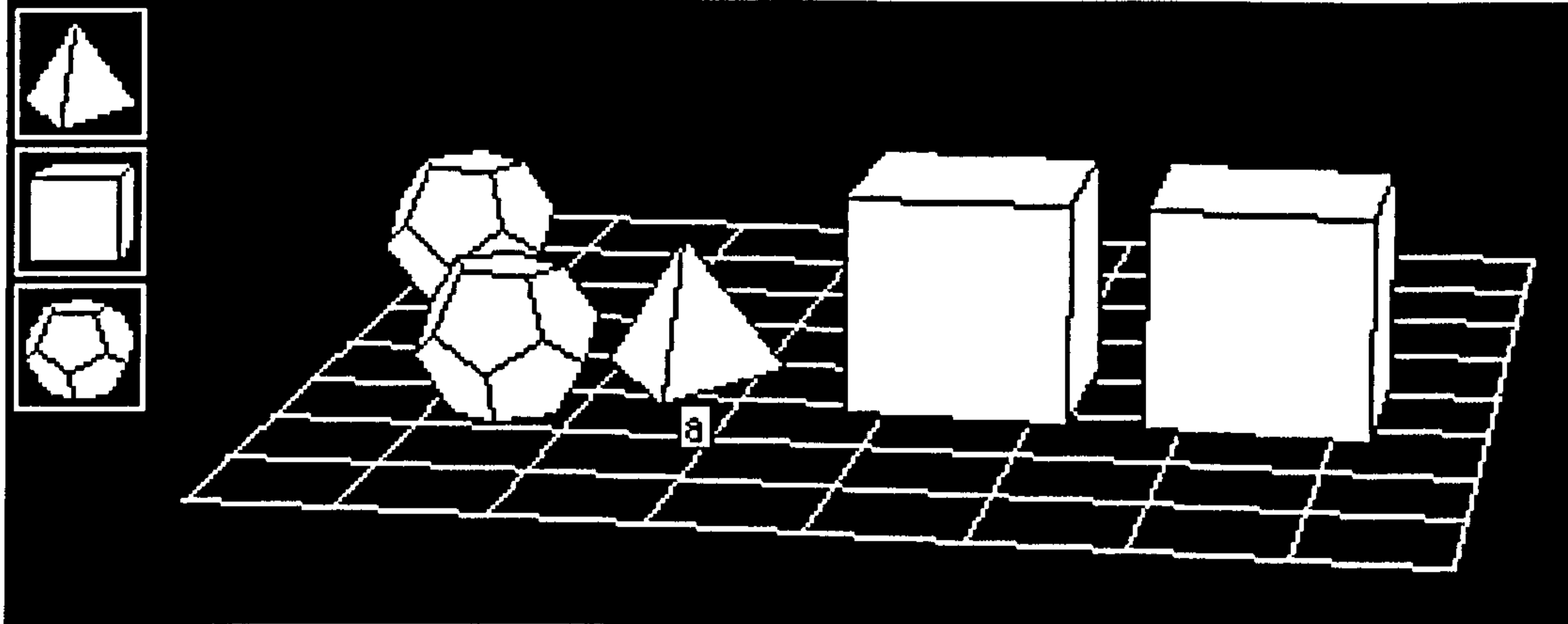
*True (but not always).*

*(Some large things are in front of 'a')  
From the diagram above this statement is also true, but if one removes the Large Dodecahedron, the premises remain true - but this statement becomes false. This statement cannot therefore be valid.*

### EXAMPLE TWO

This example of solving a syllogism using *TARSKI'S WORLD*, is in the form AEE. In this example and the last, only the current solution is shown.

**All Marathoners are Runners**  
**No Joggers are Runners**  
**No Joggers are Marathoners**



$\forall x (\text{Cube}(x) \Rightarrow \text{Large}(x))$

*(All Cubes are Large)*

*[Replacing the Marathoners with Cubes and the Runners with Large objects, and maintaining the form All As are Bs (A); the first premise looks like this]*

$\forall x (\text{Leftof}(x,a) \Rightarrow \neg \text{Large}(x))$

*(Nothing to the left of a is large)*

*[Replacing the Joggers with objects leftof 'a', and again Runners with Large objects and maintaining the form of the premise No Cs are Bs (O); the second premise looks like this.]*

$\forall x (\text{Leftof}(x,a) \Rightarrow \neg \text{Cube}(x))$

*(Nothing to the left of a is a cube)*

*[Interrelating the end terms, C & A (Cubes and Leftof), the only conclusion which cannot be made false whilst the premises remain true is this, No Cs are As (E).]*

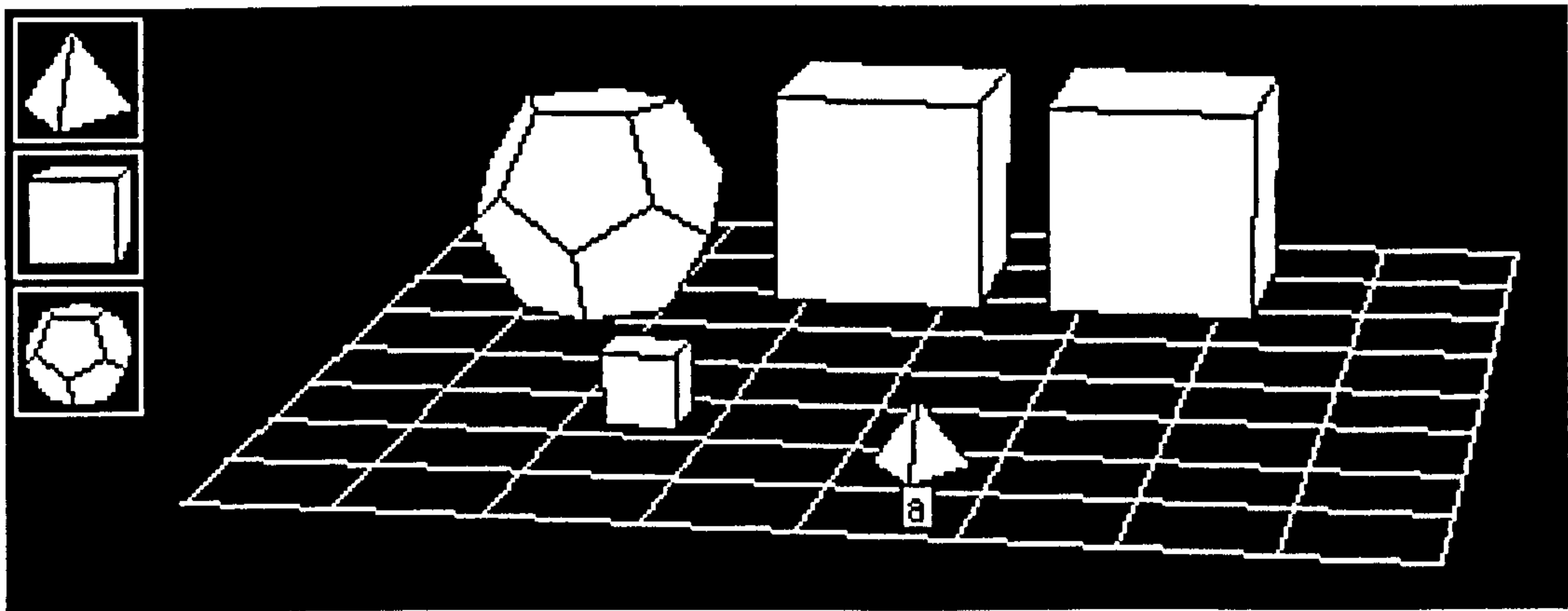
**No Joggers are Marathoners**

*[Translating the Leftof(x,a) back to Joggers and the Cube(x) back to marathoners.]*



EXAMPLE THREE.

Some Doctors are Smokers  
All Smokers are Athletes  
Some Athletes are Doctors



$\exists x (\text{Cube}(x) \wedge \text{Large}(x))$

*(Some Cubes are Large)*

*[Replacing Doctors with Cubes and Smokers with Large things, maintaining the form Some As are Bs (I); the first premise looks like this.*

$\forall x (\text{Large}(x) \Rightarrow \text{Backof}(x,a))$

*(All large things are behind 'a')*

*[Replacing Smokers with Large things again and Athletes with things behind a, whilst maintaining the form of the premise All Bs are Cs (A); the second premise looks like this.*

$\exists x (\text{Backof}(x,a) \wedge \text{Cube}(x))$

*(Some things behind 'a' are cubes).*

*Interrelating the end terms (A & C), Cubes and Backof(x,a), the only conclusion which cannot be made false whilst the premises remain true is this, Some As are Cs (I)*

Some Athletes are Doctors


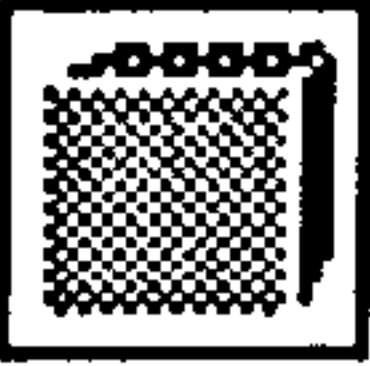

Self-test with *TARSKI'S WORLD*

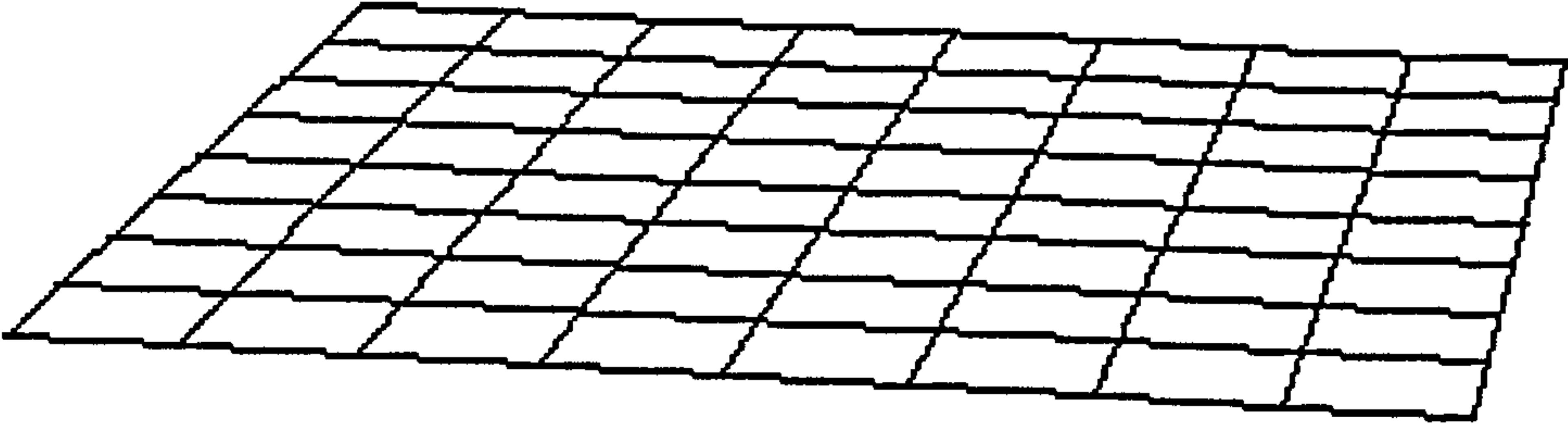
Take the premises;  
  
*All Academics are Absent minded*  
*No Computers are Absent minded*

Convert the terms into geometric relations which can be shown in the *TARSKI'S WORLD* program (remember to use unrelated terms). Use the same relations as shown in the first example *i.e.* *Academics* can be translated to *Large*, *absent-minded* can be translated to *Tetrahedron* and *Computers* can be translated to *Frontof* (*x,a*).

Write in the first box provided the FOPC equivalent, and draw in the second box a world where the premises would be true.

The form of the premise	The FOPC equivalent in <i>TARSKI'S WORLD</i>









## The Euler circle system and the syllogism.

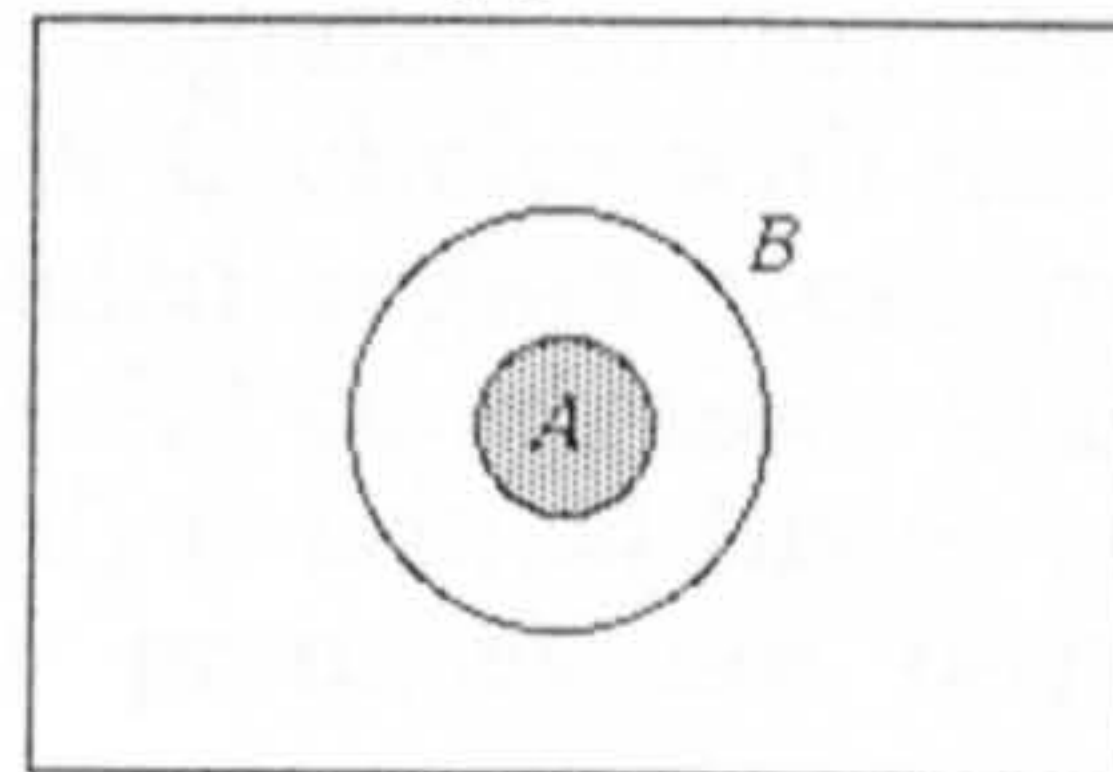
### INTRODUCTION

*Euler* diagrams can be used to represent and help solve syllogistic reasoning problems. You will remember from the introductions tutorial that there are three terms in a syllogism. Similarly in the *Euler* circle system there are three circles, one for each of the terms. Each premise in the syllogism tells us something about two of those terms and the shading conventions we use to show this information are described in detail in the examples over the page. The *Euler* circle system software which you will be using will tell you when you have constructed the diagram correctly and will tell you whether there is a valid conclusion to the premises of your syllogism. There are a number of key conventions in *Euler* which you should remember from the outset.

- (1) *Each circle represents a term.*
- (2) *Shading represents the existence of at least one individual in a particular region.*
- (3) *There are only four ways in which the circles of two terms can be interrelated.*
- (4) *The combination (or registration) of those two premise circle combinations must make as many separate regions, which still remain consistent with the premises.*

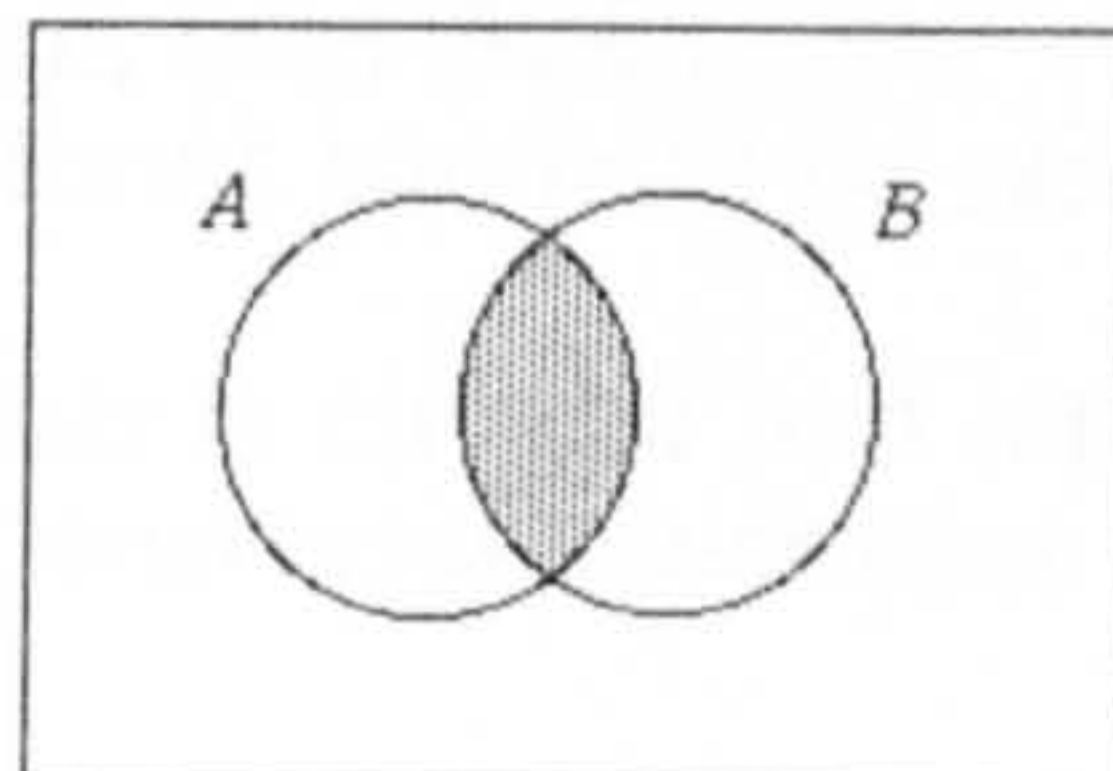
### EULER CIRCLES

There are four characteristic premise diagrams for the four possible premises which can occur in a syllogism. They are listed here below. You will find these diagrams drawn at the top of the window in the Euler Application.



All As are Bs

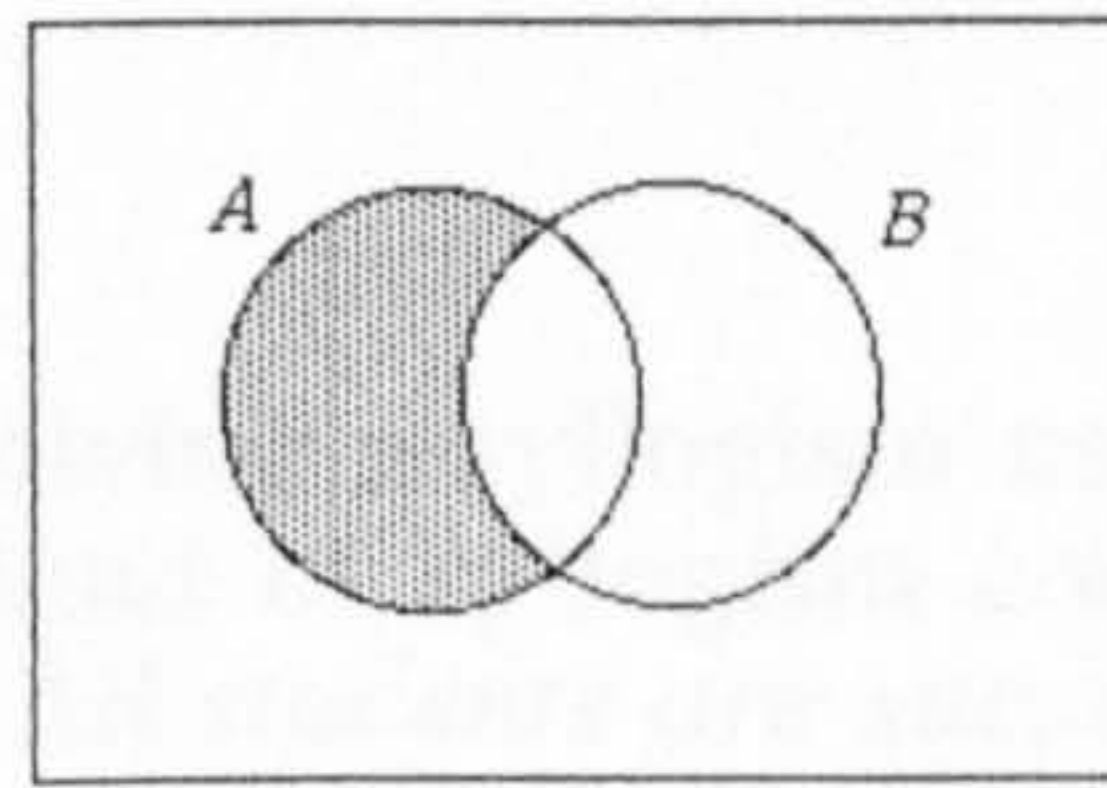
*[The A circle is contained within the B circle - indicating that All As are also Bs].*



Some As are Bs

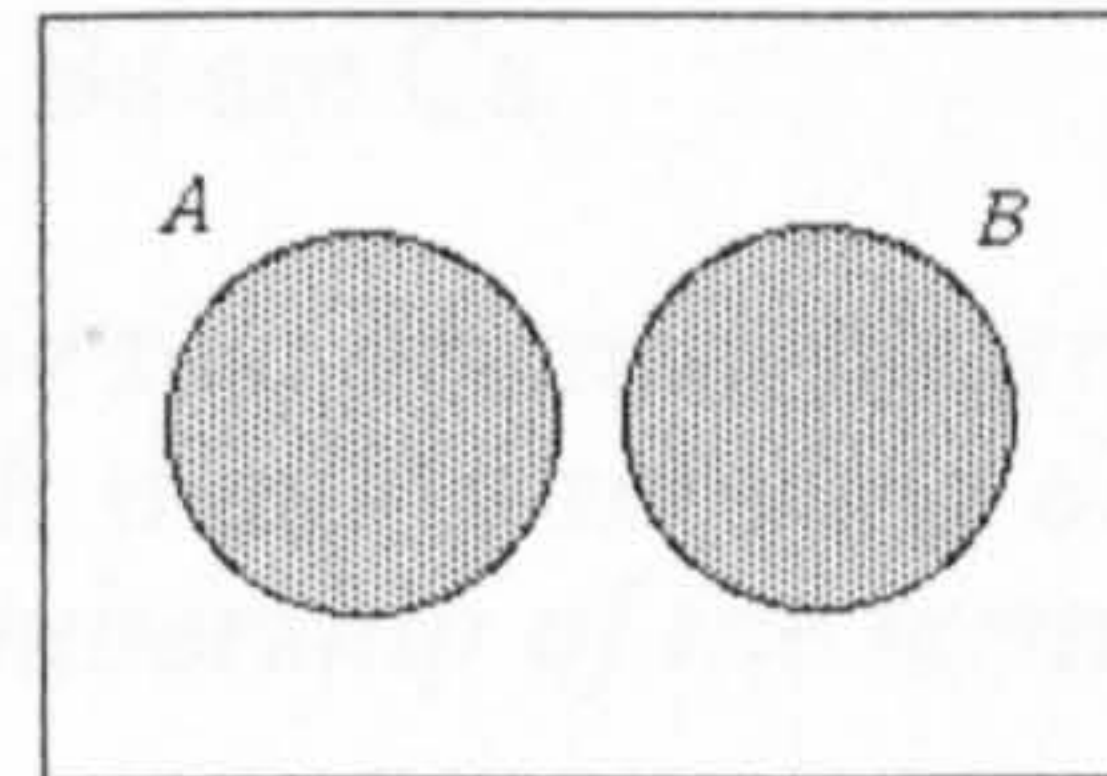
*[The shading represents individuals which actually exist, in this case some of the As are Bs].*





Some As are not Bs

*[The shaded area represents individuals which are As but not Bs, the intersection is not shaded as we do not know whether any As are Bs too].*



No As are Bs

*[There are individuals in both A and B - but not both A and B].*

1. Dynamically one can imagine the A circle expanding to fill the B circle or shrinking or moving about within the B circle. In all these cases it would still be contained within B and therefore represents All As are Bs.
2. Dynamically one can imagine pushing the A and B circles further together so the area of As which are Bs increases, as long as there is still an area where A does not completely overlap with A. Also they could be pulled apart (as long as they don't become completely detached), making less area which is A and B.
3. By Pulling A away from B, more As become not Bs, and by pushing them together less As become not Bs.
4. As long as A and B remain separated then No As will be B.

### SOLVING SYLLOGISMS

1. Convert the premises into the form which underlies them, using As Bs and Cs (remember from the introduction tutorial). Then select from the list the characteristic diagrams for each of the premises. The diagrams which are characteristic of these premises are shown in figure above.
2. Arrange the two diagrams so that the B circles are superimposed on each other. Arrange A & C circles with most regions consistent with premises.
3. If there is no shaded region from premises which remain non-intersected then there is no valid conclusion. Otherwise, that region (the one which is still not intersected) is critical (more later).
4. However if both premises are negative then, again there is no valid conclusion.
5. Describe the critical region, (e.g. A,  $\neg$ B and C).
6. Remove the B term, (e.g. A, C).
7. The A and the C terms can now be said to exist and a Some A-C conclusion made.
8. If the critical region is circular and labeled by an end term (A or C); then it is the subject of a universal conclusion (All As are Cs, All Cs are As).



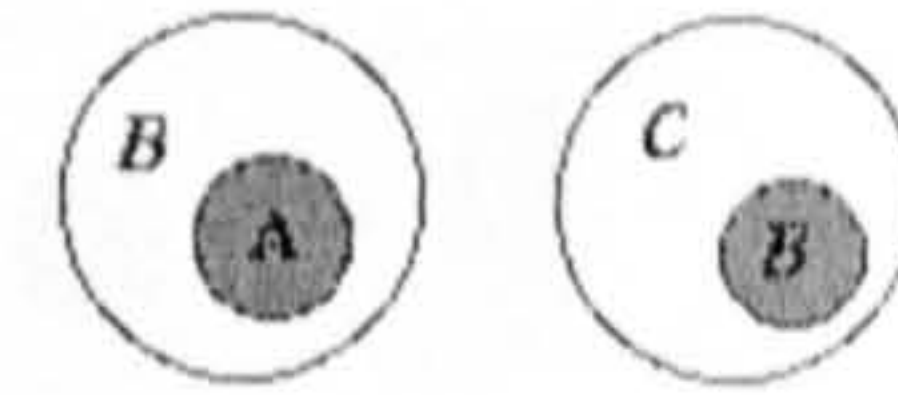
### EXAMPLE ONE

This example of solving a syllogism using the *Euler* system, is in the AAA form. You will remember that this kind of syllogism could be *All students are clever, All clever people are successful, therefore All students are successful.*

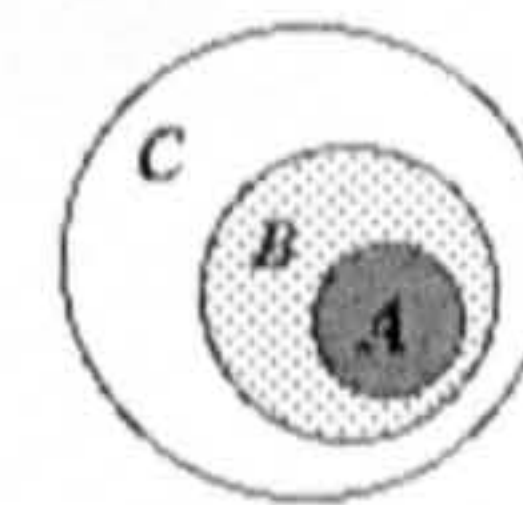
E.g  
. 1.  
[A  
AA  
]

All As are Bs  
All Bs are Cs

*[The two premise diagrams opposite with shading showing definite membership of the term].*



*[Registration of the B circles from each of the two initial premises leaves us with A inside B and B inside C].*



*Notice that the A circle remains non-intersected so is the subject of a valid conclusion. The name of the region is {A,B,C}, when the B term is removed - Some A are C*

*Because it is circular it has a universal conclusion.*

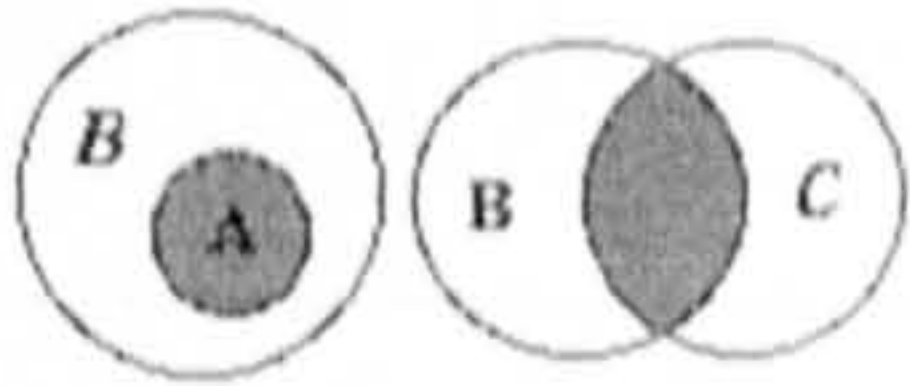
All As are Cs



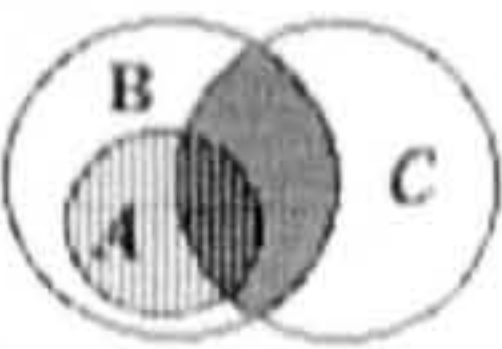
EXAMPLE TWO

In this example, there is no valid conclusion. The syllogism is of the form *EI*. A linguistic example would be *All teachers are bright, Some Bright people are beautiful*, where nothing valid can be concluded.

E.g. 2.                      All As are Bs  
[EI-]                        Some Bs are Cs



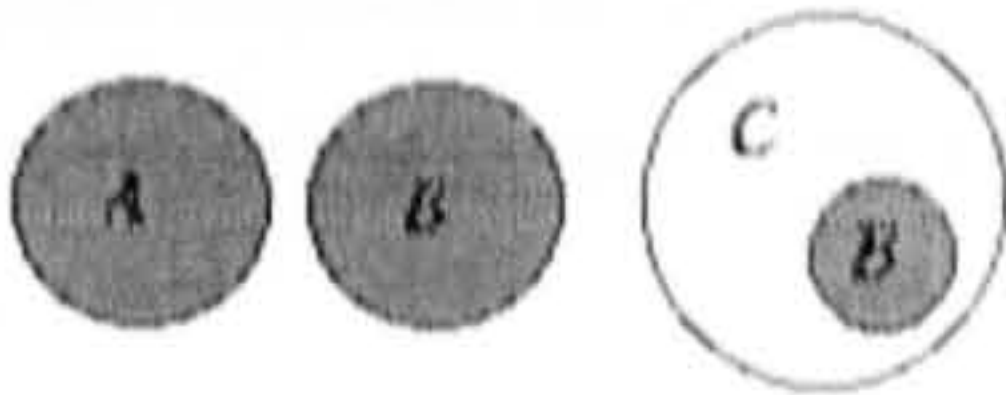
*[Registration of the B circles leaves the A intersecting the intersection of B and C. In this case there is no non-intersected area and so No valid conclusion]*



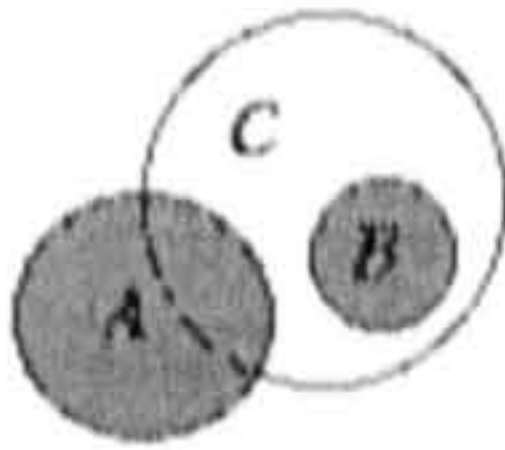
*No valid conclusion*

EXAMPLE THREE

E.g. 3.                      No As are Bs  
[OAE]                        All Bs are Cs



*[Registration of the B circles leaves us with the one non-intersected region (C,B and not A). Notice the A circle cannot intersect with the B circle because of the first premise].*



*When the B term is removed, the region is called {C,not A}*

*Some Cs are not As is derived].The region is not circular so remains an existential conclusion.*

*Some Cs are not As*





## The VENN system and Aristotelian logic.

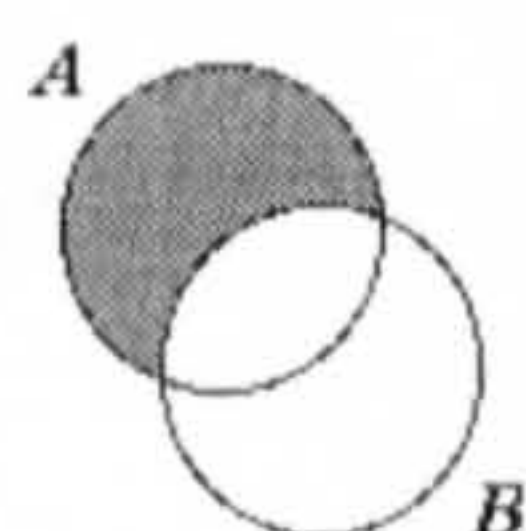
## INTRODUCTION

Venn diagrams can be used to represent and help solve syllogism problems. You will remember from the introduction tutorial that there are three terms in a syllogism. Similarly, there are three circles in the Venn diagram, one for each of the terms. Each premise in the syllogism tells us something about two of those terms and the shading conventions we use to show this information are described in detail over the page. The VENN system software also provides a number of additional features, such as the form and mood of the syllogism. You can be aware of this extra information, but you will not need it to solve the syllogisms. The Venn diagram for solving and representing syllogisms is made up of the following conventions;

- (1) Each circle represents a term.
- (2) Shading represents non-existence of a member in a particular region
- (3) The star represents existence of at least one individual, which can be either within a region or in one of two regions when the star on a line.

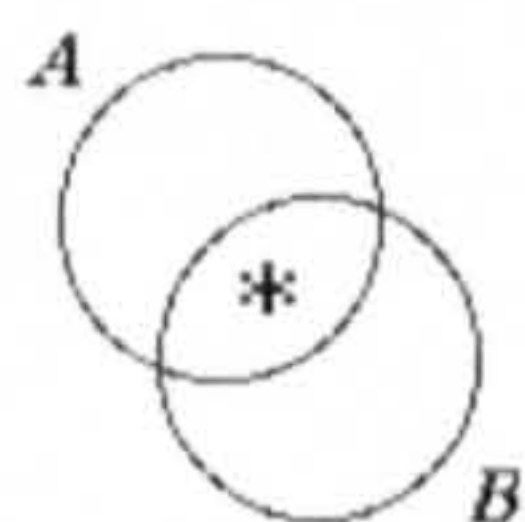
## THE SYLLOGISM

There are four premise types, (A, I, E, O), each of which has its own diagram. These four diagrams are shown below in the A-B, position. This position is the way we show the first premise of a syllogism. The second premise simply adds another circle (the C circle), which you can see in the examples over the page. Relationships between the B and C terms are shown exactly the same way as those between the A and B terms.



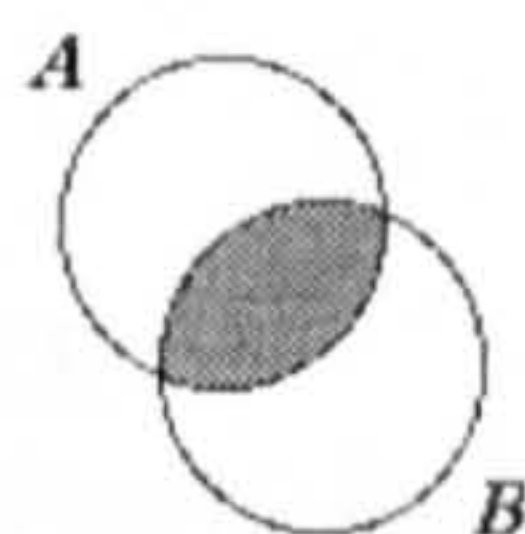
All As are Bs (A)

*The premise All As are Bs is annotated on the diagram by shading the remaining part of the A circle. There are no As which are not Bs.*



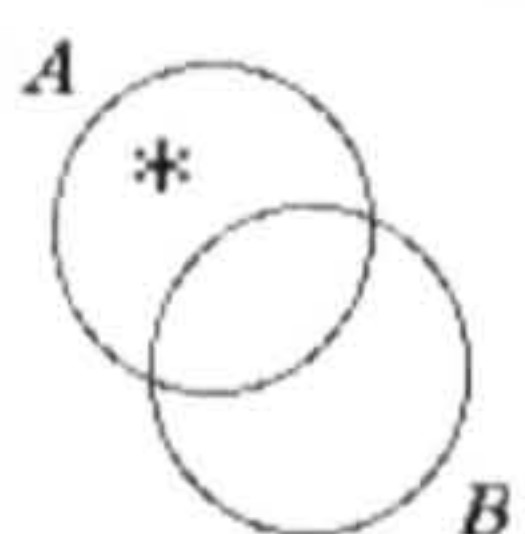
Some As are Bs (I)

*The premise Some As are Bs is shown with a star in the intersection of A and B. This signifies at least one member which is both A and B.*



No As are Bs (E)

*The premise No As are Bs is shown by shading the intersection of A and B. This indicates there are no individuals which are both A and B.*



Some As are not Bs (O)

*The premise Some As are not Bs, is shown by the star in the region of A which is not in B. Thus, there is at least one A which is not B.*

This is the first of three examples which describe Worked explanations of the use of VENN. Each example is demonstrated in stages. This example, is in the AAA form. You will remember that this kind of syllogism could be; *All students are clever, All clever people are successful, therefore All students are successful*



In the following example there is no valid conclusion. The syllogism is of the form *EI*. A linguistic example would be *All teachers are bright, Some Bright people are beautiful*, where nothing valid can be concluded.

E.g. 2 [EI-]

All teachers are Bright

All As are Bs

[Premise one from the diagram list above]

Some Bright people are beautiful

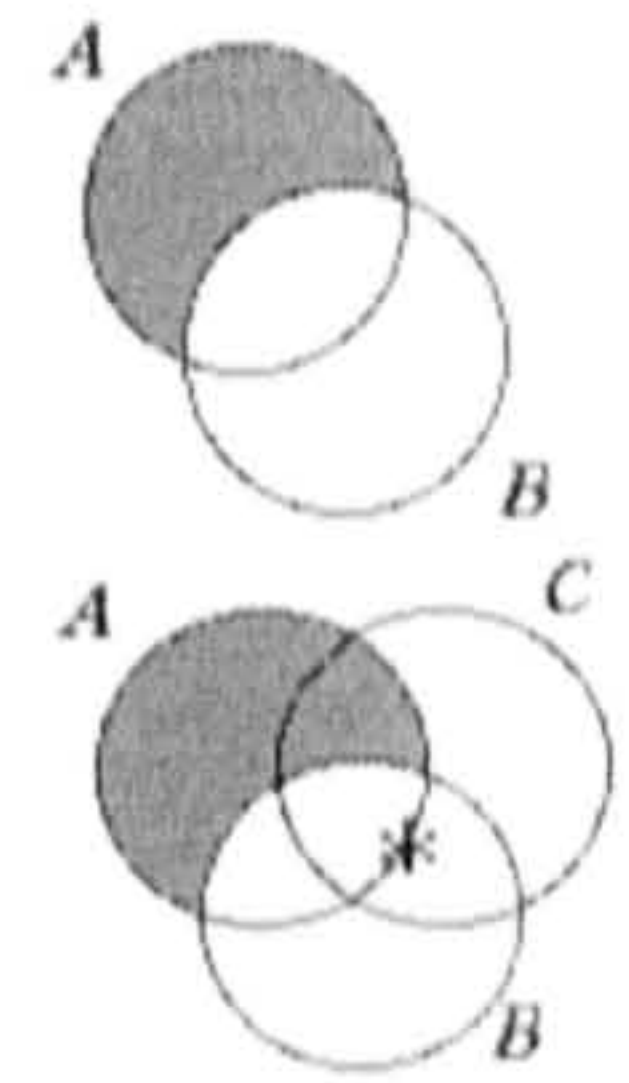
Some Bs are Cs

[The star is added in the intersection between B and C. Because it is not known whether the star is in A or not, the star remains on the line of intersection].

No valid conclusion

No valid conclusion

There is no valid conclusion as we cannot say for sure anything about A and C.



The last example is of the form *OAE*, and has a valid conclusion, *E*. A linguistic example might be, *No Politicians are honest, All honest people are wealthy, Some wealthy people are not politicians*.

E.g. 3 [OAE]

No Politicians are honest

No As are Bs

[Premise two from diagram list above]

All honest people are wealthy

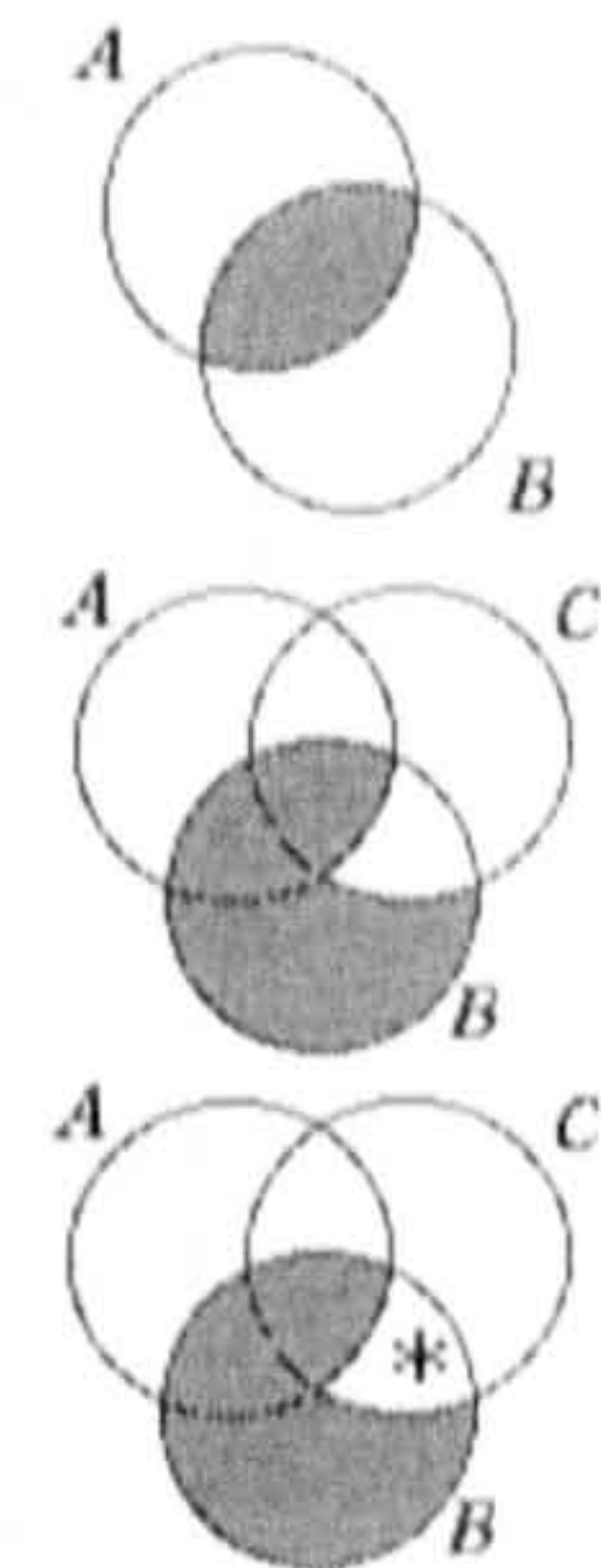
All Bs are Cs

[Diagram one from premise list, Although this time relating the B and C terms].

Some wealthy people are not politicians

Some Cs are not As

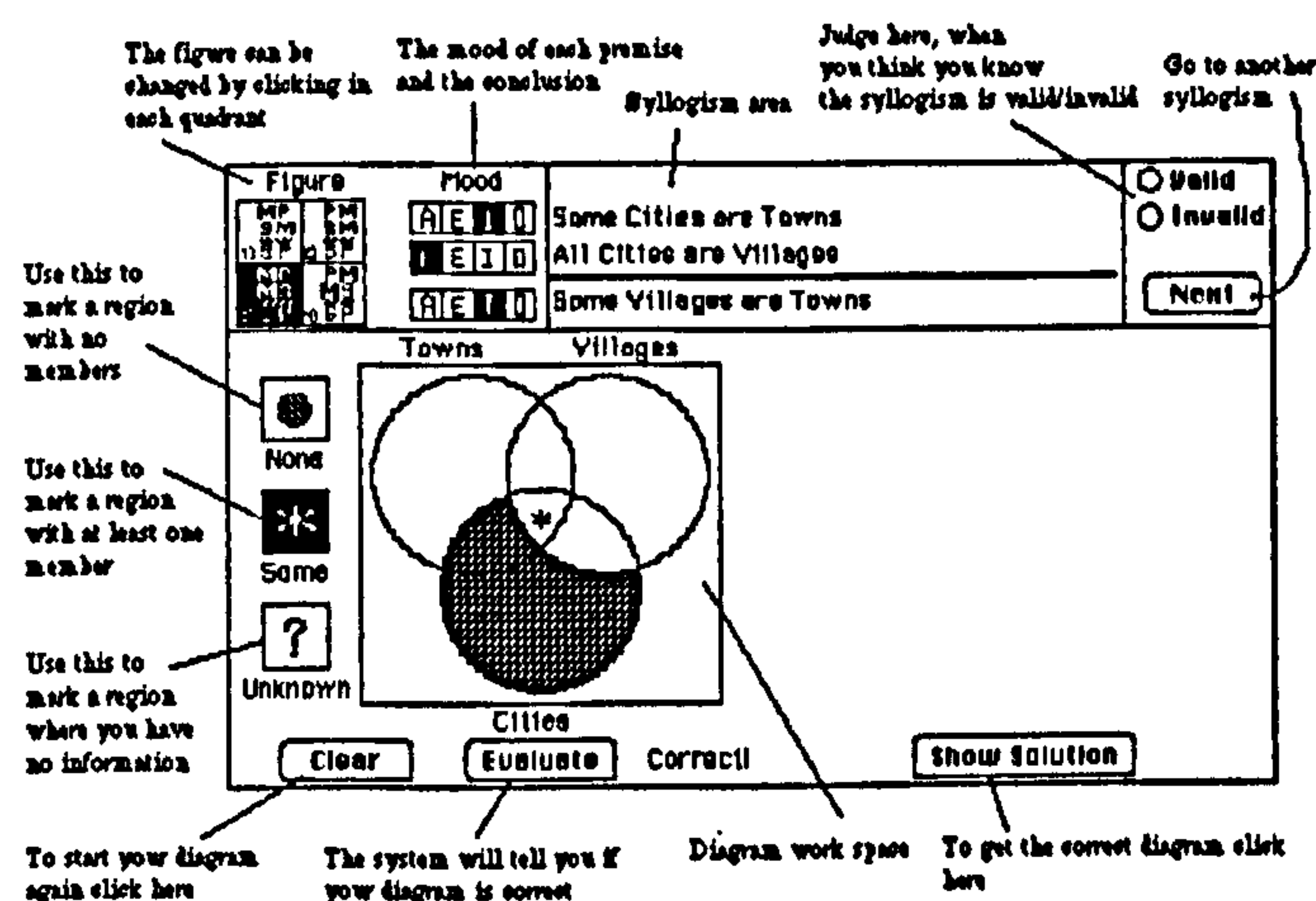
[There is only one region left of B, and since this cannot be empty there is an individual which is B and C and not A, therefore the conclusion is "Some Cs are not As"].



## THE VENN SYSTEM

The VENN system looks like the diagram below, although of course it is really twice as large as this picture. The best way to really get to know what everything does is to explore it all for a short while.





1. Click on the VENN system icon
2. Pull down the *level* menu and choose *syllogisms*
3. Read the syllogism in the top central window area. You may think it is valid or invalid, you can test this by clicking on the valid or invalid radio buttons to the left - but at this stage you may not be sure of the solution. You should then use the diagram at the lower part of the screen to work out the solution.
4. When you are happy with the solution and think you understand it, go onto the next syllogism by clicking the *Next* button.
5. Use the system in whatever way helps you to learn and remember how to solve the syllogisms using the Venn diagrams.

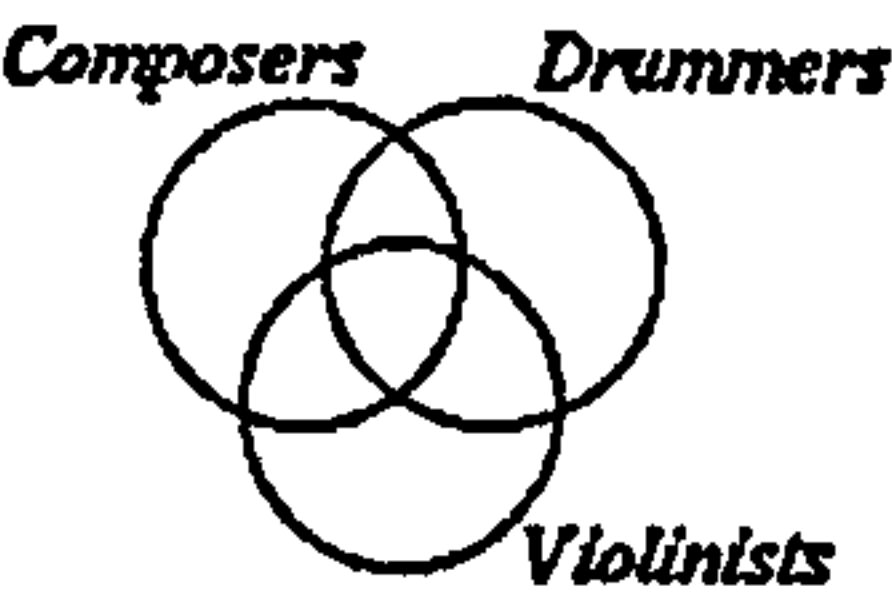
### USING THE SOFTWARE.

Take the premises; these are generated for you in the central window of the VENN interface.

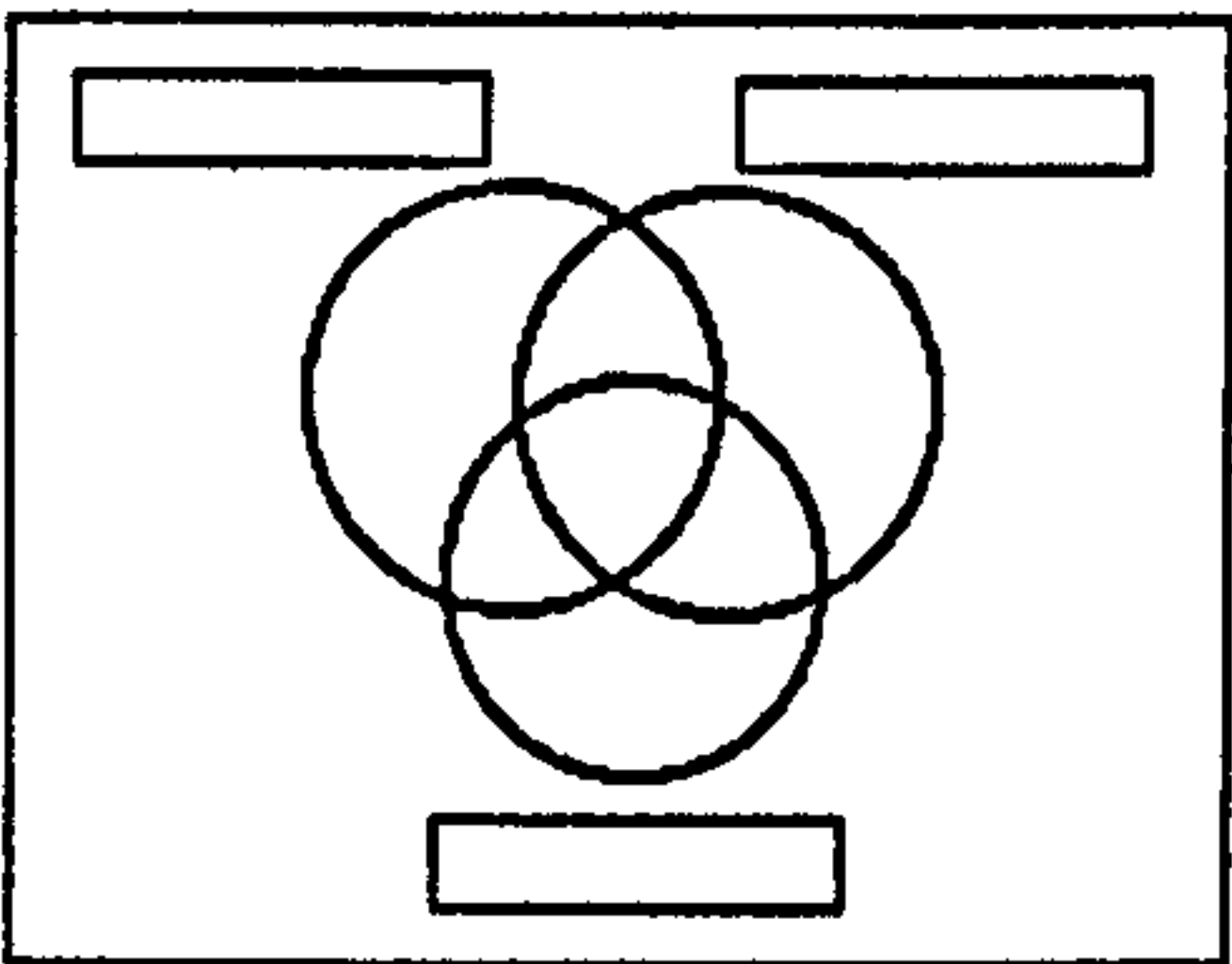
Some composers play violin  
No violinists play drums

1. Convert the premises into the form which underlie them, i.e. in this case Some As are Bs (for the first premise), and No Bs are Cs (for the second premise).
2. The diagram workspace (see VENN system interface), will already be labeled with the correct terms, you must shade them using the buttons to the left. Click on the button marked 'No' for shading, the button marked 'Some' to add a star with the mouse to the diagram and the '?' button will clear a region if you decide to 'unshade' it. Shading an area represents the impossibility of there being a member of that particular region. Adding a star somewhere to the diagram means that there is definitely a member of the region which exists. A star on a line separating regions means there is a member in one of the regions which the line separates.
3. Annotate the diagram according to these principles. (You can try this example out on the diagram grid below. For the first premise add a star in the region where there are composers and violinists, of course we don't know if the star should be within the drummers circle or not - so we put it ON the line between the two regions. For the second premise shade in the region which is common to violinists and drummers.
4. You will notice that the shaded region occupies one of the areas for which there was definitely a member (from the first premise). This star must now be in the only remaining region (the star moves into that region).
5. Read off the conclusion (it can only be between the A and the C terms in the syllogism), in this case Some As are not Cs. This can be translated into the conclusion *Some Composers don't play drums!*

*TEST.*  
Use this diagram below to practice the shading conventions, following the steps exactly as the instructions in the previous section.



Self-test with Venn
Take the premises;  All Academics are Absent-minded No Computers are Absent-minded  1. Identify the terms in the syllogism and write them into the label areas in the answer box below.  2. Fill in the diagram below with the correct shading conventions.  3. Read off the conclusion (remember it must be between the end terms, (A and C), and write it in the second box below.



Conclusion



Pre-test

In the following table you will be presented with sentences referring to three terms. The first sentence will always refer to the relationship between two terms and the second to the relationship between one term from the first sentence and a new third term. The task is to conclude any valid information which can be shown as a result of the first two sentences. In all cases the conclusion if there is any will be a relationship between a term in the first sentence and one in the second.

Here is an example:-

*All Architects are Body builders.*

*All Body builders are Crazy.*

To which you may reply (validly),

All Architects are Crazy

Now you try.... the first one is filled in for you and is the same as the example outlined above.

	<i>Example</i>	<i>Valid or Invalid</i>
1.	All Architects are Body-builders. All Body-builders are Crazy. All Architects are Crazy	V

There is no time limit, the most important thing is that you end up with what you think is the right answer... Finally, thank you for taking part in this study.

	<i>Syllogism</i>	<i>Valid or Invalid</i>
1.	Some anchovies are bitter No bitter things are chewy Some anchovies are not chewy	
2.	No pencils are pens All pens are writing tools No writing tools are pencils	
3.	No rugby players are old Some old people are fit Some fit people are not rugby players	
4.	All warlocks are witches Some magicians are warlocks All witches are magicians	
5.	All sailors are athletic No swimmers are sailors No athletes are swimmers	
6.	Some woods are hard No houses are wood Some hard things are not houses	
7.	No bullies are army soldiers Some girls are bullies Some girls are not army soldiers	
8.	All astronauts are smart Some teachers are not smart Some teachers are not astronauts	
9.	No firemen are nurses Some astronauts are nurses Some firemen are not astronauts	
10.	All business people are astute people All business people are clever All astute people are clever	
11.	All business-people are aggressive Some business-people are clever Some aggressive people are clever.	
12.	All brain-surgeons are aggressive no brain surgeons are clever No aggressive people are clever	
13.	Some buskers are talented All buskers are clever Some talented people are clever	
14.	Some marathoners are joggers No marathoners are sprinters Some joggers are not sprinters	
15.	No runners are walkers All runners are sprinters Some sprinters are not walkers	



*Examples*

Some Isotopes are not Elements  
No Atoms are Elements  
Some Atoms are Isotopes

Some Doctors are Smokers  
All smokers are Athletes  
Some Athletes are Doctors

Some Athenians are not Greeks  
Some Greeks are Men  
No Men are Athenians

All C are A  
No B are A  
No B are C

All marathoners are Runners  
No Joggers are Runners  
No Joggers are Marathoners

All Animals are Amphibians  
No Amphibians are Frogs  
No Frogs are Animals

All Athenians are Men  
All Athenians are Greeks  
Some Greeks are not Men

All Greeks are Men  
Some Greeks are not Athenians  
No Athenians are Men

Some Idealists are not Reformers  
All Idealists are Fanatics  
Some Fanatics are not Reformers

All Boats are Sloops  
Some Sloops are Ships  
No Ships are Boats

No Mountains are Hills  
Some Mountains are Volcanoes  
Some Volcanoes are not Hills

No Amphibians are Frogs  
Some Amphibians are Animals  
Some Animals are not Frogs

All Computers are Humans  
No thinkers are Humans  
Some Thinkers are not Computers

Some Trucks are not Cars  
All Trucks are Vehicles  
Some Vehicles are not Cars

Code for the Venn counting argument



**Code for the Venn counting argument**

```
#include <stdio.h>
```

```
unsigned short int vertex[8] ;
```

```
unsigned short int edge[12] ;
```

```
/* ----- */
```

```
char* binary(unsigned short int x)
```

```
{
```

```
static char result[9] ;
```

```
unsigned short int i;
```

```
unsigned short int shift ;
```

```
for (i=0;i<9;i++)
```

```
    result[i] = '\0' ;
```

```
for (i=0;i<8;i++)
```

```
{
```

```
    if (i == 0)
```

```
        shift = 1 ;
```

```
    else
```

```
        shift = 2 << (i-1) ;
```

```
    if ( ( x & shift ) == 0 )
```

```
        result[7-i] = '0' ;
```

```
    else
```

```
        result[7-i] = '1' ;
```

```
}
```

```
return (&(result[0]));
```

```
}
```

```
/* ----- */
```

```
int power(int x, int n)
```

```
{
```

```
if (n == 0)
```

```
    return (1);
```

```
return (x*power(x,n-1));
```

```
}
```

```
/* ----- */
```

```
void init_edges()
```

```
{
```

```
edge[0] = vertex[0] | vertex[1] ;
```

```
edge[1] = vertex[1] | vertex[2] ;
```



```

edge[2] = vertex[2] | vertex[3] ;
edge[3] = vertex[0] | vertex[3] ;

edge[4] = vertex[4] | vertex[5] ;
edge[5] = vertex[5] | vertex[6] ;
edge[6] = vertex[6] | vertex[7] ;
edge[7] = vertex[4] | vertex[7] ;

edge[8] = vertex[0] | vertex[4] ;
edge[9] = vertex[1] | vertex[5] ;
edge[10] = vertex[2] | vertex[6] ;
edge[11] = vertex[3] | vertex[7] ;

return;
}

/* ----- */

int count_edges()
{
int result ;
int i;

result = 0;
for (i=0;i<12;i++)
    if (edge[i] > 0)
        result++ ;

return result;
}

/* ----- */

void main(void)
{
int i, j;
unsigned short int shading;
int shaded[9];
int shade_count ;
int num_edges;
int num_unshaded ;
int total[9] ;
int sub_total, grand_total ;

vertex[0] = 1 ;
vertex[1] = 2 ;
vertex[2] = 4 ;
vertex[3] = 8 ;
vertex[4] = 16 ;
vertex[5] = 32 ;
vertex[6] = 64 ;
vertex[7] = 128 ;

init_edges();

for (i=0;i<9;i++)

```

```

{
    shaded[i] = 0;
    total[i] = 0;
}

for (i=0;i<8;i++)
    fprintf(stdout,"vertex[%2d] = %s\n",i,binary(vertex[i]));

for (i=0;i<12;i++)
    fprintf(stdout,"edge[%2d] = %s\n",i,binary(edge[i]));

fprintf(stdout,"num edges = %d\n",count_edges());

for (shading = 0;shading < 256;shading++)
{
    /*      fprintf(stdout,"shading = %s\n",binary(shading));
    */
    init_edges();

    shade_count = 0 ;
    for (i=0;i<8;i++)
    {
        if ( (shading & vertex[i]) == vertex[i])
        {
            /* the vertex is shaded */
            shade_count++ ;

            for (j=0;j<12;j++)
            {
                if ((edge[j] & vertex[i]) == vertex[i])
                    edge[j] = 0 ;
            }
        }
    }

    num_unshaded = 8 - shade_count ;
    num_edges = count_edges();
    sub_total = power(2,num_unshaded+num_edges);

    fprintf(stdout,"unshaded = %d  edges = %d sub_total =
%d\n",num_unshaded,num_edges,sub_total);

    /*
    fprintf(stdout,"shade_count = %d\n",shade_count);
    */

    shaded[shade_count]++ ;
    total[shade_count] += sub_total ;
}

grand_total = 0 ;
for (i=0;i<9;i++)
{
    grand_total += total[i];
}

```

*all possible shadings*



```
        fprintf(stdout,"shaded[%d] = %d  states = %d\n",i,shaded[i],total[i]);
    }
    fprintf(stdout,"grand_total = %d\n",grand_total);
    return;
}
```